

45131  
SEARCH REQUEST FORM

Requestor's Name: CHRIS DRABIK Serial Number: 09/445201

Date: 6/19/01 Phone: 605 1156 Art Unit: 1633  
1234

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

PLEASE SEARCH SEPARATELY THE FOLLOWING  
SEGMENTS OF SEQ ID #1 (NUCLEIC ACID ONLY)

A 8260 → 10560

B 8336 → 10608

C 10094 → 10608  
ED-

D 6036 → 6959

IF 100% HITS SHOW UP FOR A, B, & C (i.e. ALL 3  
MUST BE 100%)  
PLEASE USE THOSE HITS TO SEARCH AGAINST  
THE ENTIRE SEQ ID #1 SEQUENCE.

THANKS

- CHRIS

9 HRS

6/20/01  
6/22/01  
STAFF USE ONLY

Date completed: 6/22/01	Search Site	Vendors
Searcher: _____	<input checked="" type="checkbox"/> STIC	_____ IG
Terminal time: _____	<input checked="" type="checkbox"/> CM-1	_____ STN
Elapsed time: _____	_____ Pre-S	_____ Dialog
CPU time: _____	Type of Search	_____ APS
Total time: _____	<input checked="" type="checkbox"/> A.A. Sequence	_____ Geninfo
Number of Searches: _____	_____ A.A. Sequence	_____ SDC
Number of Databases: _____	_____ Structure	_____ DARC/Questel
	_____ Bibliographic	<input checked="" type="checkbox"/> Other



11	42.2	1.9	14507	21
----	------	-----	-------	----

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2273	100.0		12845	20	V84274	Mouse endothelial
C 2	55.4	2.4		5513	21	C63529	Slit protein codin
C 3	46	2.0		4064	21	A07587	Mouse piwi gene, d
4	45.6	2.0		51259	18	X83007	Partial mouse WRN
C 5	45	2.0		2362	21	Z50462	Human tissue trans
C 6	44	1.9		1110	19	V29343	Calcium ion channe
7	43.6	1.9		810	20	Z16495	Human gene express
8	43.2	1.9		11901	20	Z02998	Human IL-1ra BAC c
C 9	42.6	1.9		8966	20	Z09581	Human Apo A1 genom
10	42.2	1.9		14507	18	T73568	Expression augment
11	42.2	1.9		14507	21	Z88869	Chinese hamster 2A

## ALIGNMENTS

## RESULT 1

V84274  
ID V84274 standard; DNA; 12845 BP.

V84274:

12-APR-1999 (first entry)

Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.

Flk-1; endothelial growth factor receptor-2; VEGF; transcription; endothellum; enhancer; vascular disease; angiogenesis; cancer; diabetic retinopathy; rheumatoid arthritis; wound healing; vulnary; atherosclerosis; tumour; neuronal disorder; therapy; diagnosis; mouse; ss.

AA  
MS  
SS

XX	Key	Location/Qualifiers
EH		

FT  
FT  
PROMOTER

LETTERS

5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528

333

$$\begin{array}{c} E \rightarrow E \\ \Gamma \quad \Gamma \end{array}$$
protein\_L  
F<sup>0</sup>  
F<sup>T</sup>

LE

FT protein\_bind 10166..10179  
 FT /tag= e  
 FT /note= "GATA consensus binding site"  
 FT protein\_bind 10187..10197  
 FT /tag= f  
 FT /note= "AP1 consensus binding site"  
 FT protein\_bind 10252..10261  
 FT /tag= g  
 FT /note= "NFkB consensus binding site"  
 FT protein\_bind 10273..10283  
 FT /tag= h  
 FT /note= "AP1 consensus binding site"  
 FT protein\_bind 10301..10309  
 FT /tag= i  
 FT /note= "STAT consensus binding site"  
 FT protein\_bind 10356..10372  
 FT /tag= j  
 FT /note= "ETS-1/GATA consensus binding site"  
 FT protein\_bind 10395..10404  
 FT /tag= k  
 FT /note= "ETS-1 consensus binding site"  
 FT protein\_bind 10453..10477  
 FT /tag= l  
 FT /note= "SCL/TAL-1 Ets-1 consensus binding site"  
 FT exon 10643..10726  
 FT /tag= m  
 FT /number= 2  
 FT intron 10727..12673  
 FT /tag= n  
 FT /number= 2  
 FT exon 12674..12845  
 FT /tag= o  
 FT /number= 3  
 FT XX  
 FT XX W09855638-A1.  
 FT XX  
 FT XX  
 FT PD 10-DEC-1998.  
 FT XX  
 FT PF 03-JUN-1998; 98WO-EP03318.  
 FT XX  
 FT PR 03-JUN-1997; 97EP-0108959.  
 FT XX  
 FT XX  
 FT PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 FT XX  
 FT XX Breier G, Risau W, Roenicke V;  
 FT XX  
 FT DR WPI; 1999-059915/05.  
 FT XX  
 FT XX  
 FT PT New recombinant DNA containing heterologous sequence controlled by  
 PT endothelial cell specific regulator - from the Flk-1 gene, used to  
 PT treat, prevent or diagnose vascular disease, tumours, also to screen  
 PT for transcriptional regulators  
 FT  
 FT Claim 3; Fig 1; 107pp; English.  
 FT  
 FT  
 CC This 12.8 kb DNA sequence spans the region from -6.65 kb relative to  
 CC the transcriptional start site to +6.15 (located in the third exon)  
 CC of the murine endothelial growth factor (VEGF) receptor-2 gene  
 CC Flk-1. This portion of the Flk-1 gene was isolated from the liver  
 CC of mouse strain 129/SvJ by screening a phage library of liver  
 CC genomic DNA in vector lambda FIXII. Sequences within the 5'-flanking  
 CC region of the gene, in combination with sequences located within the  
 CC first intron, specifically and reproducibly target expression of  
 CC heterologous DNA to angioblasts during early stages of vascular  
 CC development and also to the vasculature of postnatal mice. The  
 CC regulatory sequences of the first intron also function as an  
 CC autonomous endothelium-specific enhancer when fused to a  
 CC heterologous promoter (e.g. HSV-TK promoter). This Flk-1 intron  
 CC enhancer contains several potential binding sites for transcription  
 CC factors of the Ets and GATA families. The invention provides a  
 CC new claimed recombinant DNA (I) comprising at least one regulatory  
 CC sequence from an intron of the Flk-1 gene, or its homologue, to  
 CC control expression in endothelial cells, in vivo, linked to (b) a

CC heterologous DNA. (I) is used to produce transgenic animals, and  
 CC these, or transformed cells, are used to identify agents (A),  
 CC potential pharmaceuticals, that suppress, activate or enhance  
 CC transcription of genes in endothelial cells. (I), vectors and (A)  
 CC are used to direct or prevent (for antisense sequences) expression  
 CC of genes specifically in endothelial cells, e.g. for treating  
 CC angiogenesis, cancer, diabetic retinopathy, rheumatoid arthritis  
 CC etc., and in wound healing, particularly for treating vascular  
 CC disease (particularly proliferation of smooth muscle cells,  
 CC specifically atherosclerosis), tumours and neuronal disorders.  
 CC They may also be used to induce vascular disease in the transgenic  
 CC animals or diagnostically, particularly for studying (mal)function,  
 CC interactions and unregulated expression of endothelial products.  
 CC (I) provide specific modulation of gene expression in endothelial  
 CC cells, at all stages of development.  
 XX  
 SQ Sequence 12845 BP; 3321 A; 2893 C; 2863 G; 3723 T; 45 other;  
 Query Match 100.0%; Score 2273; DB 20; Length 12845;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tctcgagacagcgctgtgttttttctctgtgttttgcgagcggaagataacagtg 60  
 DB 8336 tctcgagacagcgctgtgttttttctctgtgttttgcgagcggaagataacagtg 8395  
 QY 61 cacaataataatgtgttgcgatttgaacataaaaggcgtttattgtatagta 120  
 DB 8396 cacaataataatgtgttgcgatttgaacataaaaggcgtttattgtatagta 8455  
 QY 121 gcatactactcttcgagtcgaatgagctgtctaaagaacagacccaactgcga 180  
 DB 8456 gcatactactcttcgagtcgaatgagctgtctaaagaacagacccaactgcga 8515  
 QY 181 tgaataatgaatgaggtttaataaaggcgatggatgagcattgactgataaactc 240  
 DB 8516 tgaataatgaatgaggtttaataaaggcgatggatgagcattgactgataaactc 8575  
 QY 241 cagttattgataacctcattgactggttgcgagacatgtattggtatggggcgc 300  
 DB 8576 cagttattgataacctcattgactggttgcgagacatgtattggtatggggcgc 8635  
 QY 301 ctttaagatgagcatagcgaacgctgcactcctaaagaacatctatggtatgta 360  
 DB 8636 ctttaagatgagcatagcgaacgctgcactcctaaagaacatctatggtatgta 8695  
 QY 361 ttacagacagcttgagaagctcttagtgcctgcggtgtagatcagcgtagagcgc 420  
 DB 8696 ttacagacagcttgagaagctcttagtgcctgcggtgtagatcagcgtagagcgc 8755  
 QY 421 gaggcctgcgctcctctgcactgaagaataaaggccatttactgtggtgtag 480  
 DB 8756 gaggcctgcgctcctctgcactgaagaataaaggccatttactgtggtgtag 8815  
 QY 481 tgggcgagcttgagcaggttactactatactttctcacacatcctgcgactaata 540  
 DB 8816 tgggcgagcttgagcaggttactactatactttctcacacatcctgcgactaata 8875  
 QY 541 gttcatcagatgagcgtatccagtgattgttgcaggttaagtgttcagtcagtta 600  
 DB 8876 gttcatcagatgagcgtatccagtgattgttgcaggttaagtgttcagtcagtta 8935  
 QY 601 gaattacttatcaaacaaatgttttctcatttctcgtcttctctcaaacaaagta 660  
 DB 8936 gaattacttatcaaacaaatgttttctcatttctcgtcttctctcaaacaaagta 8995  
 QY 661 ttccattatgaagcgttttgaagacattttaaactgcttgcctatgttaggacagt 720  
 DB 8996 ttccattatgaagcgttttgaagacattttaaactgcttgcctatgttaggacagt 9055  
 QY 721 gacttattcatattgacaaatattatgccgatttaattgaatgactaccagctctat 780





PT diagnostic applications -  
XX  
PS Claim 19; Page 180-185; 201pp; English.  
XX  
CC This sequence encodes the mouse piwi family protein, designated  
CC miwi. The piwi family nucleic acids and polypeptides are used in gene  
CC therapy of diseases such as cancer and also in various research and  
CC diagnostic applications. The sequences can also be used to treat  
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.  
XX  
SQ Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other:

[illegible]

XX  
XX x83007;  
XX  
XX  
DT 31-AUG-1999 (first entry)  
XX  
DE partial mouse WRN genomic sequence #3

XX	30-DEC-1996;	96WO-US20785.
PF		
XX	12-APR-1996;	96US-0632175.
XX	29-DEC-1995;	95US-0009409.
PR	29-DEC-1995;	95US-0580539.
PR	30-JAN-1996;	96US-0010835.
PR	30-JAN-1996;	96US-0594242.
XX		
XX	(DARW-)	DARWIN MOLECULAR CORP.
PA	(OSHI-)	OSHIMA J.
PA		

XX	Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;
PI	WPI; 1997-363671/33.
XX	
XX	
XX	
PT	Isolated nucleic acid molecule encoding the WRN gene product -
PT	useful for detection and treatment of Werner's syndrome, and related
PT	diseases
XX	
XX	Claim 1; Fig 7; 153pp; English.
PS	
XX	This sequence represents a fragment of the genomic sequence containing
CC	the coding region for the mouse WRN gene (X83004). The corresponding
CC	human gene (X83001) encodes a protein related to Werner's syndrome.
CC	The products can be used for the detection and treatment of Werner's
CC	syndrome (WS), an autosomal recessive disorder with a complex phenot,
CC	as well as related diseases.
CC	
XX	

SQ Sequence 51259 BP; 14533 A; 9635 C; 10266 G; 16825 T; 0 other;

	Query Match	2.0%	Score 45.6;	DB 18;	Length 51259;
	Best Local Similarity	62.1%;	Pred. No. 0.069;	Mismatches	0;
	Matches	72;	Conservative	Indels	Gaps
QY 1292	tattaattagttattcttctcttcgttgacagagatcctaactggtggcccgaggctagtct	1351			
Db 23743	ttttagtttttagtttacttttttttagacagggtctcaactggtagctggggacaagct	23802			
QY 1352	canaactggcggtccaatttgtctcaactcataagaatgctgggcttcagggtgtgtgc	1407			
Db 23803	ccaacctgtgtccctttctcaccctctcgaagtgtgggatcacagagcgtgtgc	23858			

RESULT	5	
Z50462/c		
ID	Z50462 standard; mRNA; 2362 BP.	
XX		
XX		
XX	Z50462;	
XX		
XX	23-MAY-2000 (first entry)	
XX		
XX	Human tissue transglutaminase (tTGase) homologue-3 encoding mRNA.	
XX		
XX	Tissue transglutaminase; tTGase; human; homologue; myeloid cell; LTR-HPC;	
KW	long-term repopulating haematopoietic progenitor cell; calcium dependent;	
KW	apoptosis; cell adhesion; metastasis; extracellular matrix assembly;	
KW	ECM; marker; allogeneic transplant; gene therapy; ss.	
XX		
XX	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	CDS	133..1182
FT		/*tag= a
FT		/product= "Human tissue transglutaminase homologue-3"
FT		/note= "GTP binding calcium dependent enzyme"
XX		
XX	WC200006766-A1.	
XX		
XX		
PD	10-FEB-2000.	
XX		
XX	28-JUL-1999;	99WO-US17096.
XX		
XX	29-JUL-1998;	98US-0123970.
XX		
XX		
PA	(NBL-) NEW YORK BLOOD CENT INC.	

CC is an mRNA encoding a homologue of human tGase, expressed in LTR-HPC.  
XX  
SQ Sequence 2362 BP; 616 A; 593 C; 640 G; 513 T; 0 other;

	Query Match	2.0%;	Score 45;	DB 21;	Length 2362;
	Best Local Similarity	70.6%;	Pred. No. 0.019;		
	Matches 60;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;
QY 1280	atgcctgcctgctat	aaattagttattctctctctgagacagagtcactcgtgtggc	1339		
Db 1787	ATGCCTGCTAATTTTATTTTATTTTATTTTATTTTGTAGAGACAGGCTCTACTATATTGC	1728			
QY 1340	ccaggctagctc	aaacttgcggctc	1364		
Db 1727	CCAGGCTGGTCTCAAACTCCGGGCC	1703			

RESULT	6
V29343/c	
ID	V29343 standard; DNA; 1110 BP.
XX	
AC	V29343;
XX	
XX	31-JUL-1998 (first entry)
XX	
DE	Calcium ion channel alpha subunit exon 1
XX	
KW	Calcium ion channel alpha subunit; human
KW	familial hemiplegic migraine; FHM; EA-2; 1
KW	exon; intron; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	intron
FT	Location/Qualifiers
FT	1..201
FT	/tag= a
FT	/number= 14
FT	/note= "partial sequence"
FT	202...274
FT	/tag= b
FT	/number= 15
FT	275..1110
FT	/tag= c
FT	/number= 15
FT	/note= "partial sequence"
FT	

CC episodic neurological disorders, specifically migraine, FHM or EA-2, but  
 CC also epilepsy. The isolated or a recombinant nucleic acid can also be  
 CC used to distinguish between alleles of the corresponding gene. Cells and  
 CC animals containing recombinant expression vectors comprising the nucleic  
 CC acid can be useful in study, development and treatment of migraine, FHM,  
 CC EA-2 and epilepsy. Proteins or peptides encoded by the nucleic acid and  
 CC natural or synthetic antibodies against the proteins can be used to  
 CC diagnose FHM, EA-2, migraine and other neurological conditions associated  
 CC with cation channel dysfunction.  
 XX  
 SQ Sequence 1110 BP; 306 A; 287 C; 275 G; 239 T; 3 other;

Query Match 1.9%; Score 44; DB 19; Length 1110;  
 Best Local Similarity 71.8%; Pred. No. 0.025;  
 Matches 56; Conservative 1; Mismatches 21; Indels 0; Gaps 0;  
 QY 1287 ccgtctattattagttattctctcttgagacagagtcactgtgtgcccagct 1346  
 ||||| ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||  
 Db 939 CCAGCTAATTAATTAATTAATTAATTTTTCAGAGACGGAGTCTCACTAATGTTGCCAGACT 880  
 QY 1347 agtctcaaacctgcgtc 1364  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||  
 Db 879 TGCTCAAACTCTGGCC 862

RESULT 7  
 Z16495 126495 standard; cDNA; 810 BP.  
 XX  
 AC Z16495;  
 XX  
 DT 12-OCT-1999 (first entry)  
 XX  
 DE Human gene expression product cDNA sequence SEQ ID NO:3965.  
 DE  
 KW Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9938972-A2.  
 XX  
 PD 05-AUG-1999.  
 XX  
 PF 28-JAN-1999; 99WO-US01619.  
 XX  
 PR 03-APR-1998; 98US-0080666.  
 PR 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 XX  
 DR WPI; 1999-494092/41.  
 XX

Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 PS  
 PS Claim 1; Page 1883; 2479pp; English.  
 XX

CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in Z12532 to Z17779. Also described is a  
 CC method of detecting differentially expressed genes correlated with the

CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in Z12532 to Z17779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX  
 SQ Sequence 810 BP; 256 A; 145 C; 148 G; 253 T; 8 other;

Query Match 1.9%; Score 43.6; DB 20; Length 810;  
 Best Local Similarity 64.3%; Pred. No. 0.027;  
 Matches 81; Conservative 0; Mismatches 44; Indels 1; Gaps 1;  
 QY 1293 attaatgtattctctctcttgagacagagtcactgtgtgcccagctagtctc 1352  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||  
 Db 418 ataacacaactttctgttttttggagacagagtcactctccaccaggctggtc 477  
 QY 1353 aaactgctggtccattgtctcactcatcagaatgctgggtctccaggtgtgtgcaccac 1412  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||  
 Db 478 aaactcccgacotcaggcgatc-cacctcaagtgctgggattacagcatgagccctgc 536  
 QY 1413 actagg 1418  
 |||||  
 Db 537 gccagg 542

RESULT 8  
 X02998  
 ID X02998 standard; DNA; 11901 BP.  
 XX  
 AC X02998;  
 XX  
 DT 22-JUN-1999 (first entry)  
 XX  
 DE Human IL-1ra BAC contiguous DNA sequence 43.  
 XX  
 KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;  
 KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;  
 KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;  
 KW growth factors; treatment; IL-1 receptor complex; BAC; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906426-A1.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 03-AUG-1998; 98WO-US16102.  
 XX  
 PR 02-JUL-1998; 98US-0091650.  
 PR 04-AUG-1997; 97US-0054646.  
 XX  
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.  
 XX  
 PI Pan Y;  
 XX  
 DR WPI; 1999-153692/13.  
 XX  
 PT New isolated nucleic acid encoding the new human cytokine Tango-77 -  
 PT used to inhibit inflammation and to screen for specific modulators



```

FT      /*tag= h
FT      /note= "the EASE sequence is preferably ligated
FT      to DNA comprising nucleotides 14290-14507
FT      (Claim 3)"
XX      WO9725420-A1.
XX      PD
XX      17-JUL-1997.
XX      XX
XX      PN
XX      10-JAN-1997; 97WO-US00483.
XX      PF
XX      11-JAN-1996; 96US-0586509.
XX      PR
XX      PA (IMMV ) IMMUNEX CORP.
XX      PI
XX      Lee C, Morris AE, Thomas JN;
XX      WPI; 1997-372861/34.
XX      DR
XX      Expression augmenting sequence elements - used in vectors for
XX      PT expressing recombinant proteins at high levels in shorter periods of
XX      PT time
XX      PS
XX      Claim 1; Page 23-30; 36pp; English.
XX      CC
XX      A 14507 bp DNA sequence (773568) comprises a novel transcription
XX      CC regulatory element, expression augmenting sequence element (EASE),
XX      CC that facilitates high expression of recombinant proteins in
XX      CC mammalian host cells. It was identified by cloning the integration
XX      CC site of a unique expression cassette encoding recombinant dimeric
XX      CC tumour necrosis factor receptor-immunoglobulin Fc fusion protein
XX      CC from genomic DNA of a CHO 2A5-3 cell line expressing this protein
XX      CC at a high level. Expression vectors incorporating an EASE show
XX      CC a 2- to 8-fold improvement of recombinant protein expression levels
XX      CC in CHO host cells. The EASE sequence also facilitates high
XX      CC expression in shorter periods of time.
XX      XX
XX      Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T; 0 other;

Query Match 1.9%; Score 42.2; DB 18; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.32;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1288 ctgctattaattagttattctctctgtgagacagagtcactgtgtgcccaggcta 1347
Db 146 cagccaagatttgttggttgtttcccgagaaaggttctctgtgagccctgctg 205

QY 1348 gtctcaact-tgcggtccattgtctcactcatcagaatgctgggtccagggtgtg 1406
Db 206 tcagggaactcagagatccaccgtcctctctctgagtgctgggattaaagggtgtg 265

QY 1407 caccacactaggtagctgcggttttaa 1433
Db 266 ctgactacaggaagctgtgttttta 292

RESULT 11
ID Z88869
XX Z88869 standard; DNA; 14507 BP.
AC Z88869;
XX Z88869;
XX 25-MAY-2000 (first entry)
DT Chinese hamster 2A5-3 lambda EASE DNA.
DE EASE; expression augmenting sequence element; chinese hamster; ds.
XX Crictetus sp.
XX US6027915-A.
XX

```

```

PD 22-FEB-2000.
XX 13-JAN-1997; 97US-0785150.
XX 11-JAN-1996; 96US-0586509.
XX (IMMV ) IMMUNEX CORP.
XX Morris AE, Thomas JN, Lee C;
XX WPI; 2000-194831/17.
XX Improved expression of recombinant proteins in stable cell pools of
XX PT mammalian cells comprises transforming cells with a vector containing
XX PT an expression augmenting sequence element -
XX PS
XX Claim 1; Column 17-30; 18pp; English.
XX CC
XX This invention describes a novel expression augmenting sequence element
XX CC (EASE) (I) isolated from Chinese hamster ovary cells. (I) facilitates
XX CC high expression of recombinant proteins in mammalian host cells in a
XX CC short time period. (I) is used for improving expression of recombinant
XX CC proteins 2-8 fold in stable cell pools when it is present in an
XX CC expression vector. (I) improves recombinant protein expression in
XX CC mammalian cells and facilitates the use of mammalian cells in
XX CC recombinant protein production.
XX XX
XX Sequence 14507 BP; 4225 A; 2548 C; 3136 G; 4598 T; 0 other;

Query Match 1.9%; Score 42.2; DB 21; Length 14507;
Best Local Similarity 59.9%; Pred. No. 0.32;
Matches 88; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1288 ctgctattaattagttattctctctgtgagacagagtcactgtgtgcccaggcta 1347
Db 146 cagccaagatttgttggttgtttcccgagaaaggttctctgtgagccctgctg 205

QY 1348 gtctcaact-tgcggtccattgtctcactcatcagaatgctgggtccagggtgtg 1406
Db 206 tcagggaactcagagatccaccgtcctctctctgagtgctgggattaaagggtgtg 265

QY 1407 caccacactaggtagctgcggttttaa 1433
Db 266 ctgactacaggaagctgtgttttta 292

RESULT 12
X37433/C
ID X37433 standard; cDNA; 687 BP.
XX X37433;
AC X37433;
XX 06-JUL-1999 (first entry)
DT Human secreted protein cDNA fragment containing gene 65.
DE
XX Human; secreted protein; prevention; treatment; protein therapy;
XX KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX KW developmental abnormality; foetal deficiency; blood disorder; lymphoma;
XX KW leukemia; immune system disorder; autoimmune disease; hepatic disease;
XX KW renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;
XX KW Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;
XX KW arthritis; psoriasis; digestive; endocrine; infection; ss.
XX Homo sapiens.
XX WO9909155-A1.
XX 25-FEB-1999.
XX 18-AUG-1998; 98WO-US17044.
XX

```



XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX WO9839448-A2.

XX 11-SEP-1998.

XX 06-MAR-1998; 98WO-US04493.

XX 03-OCT-1997; 97US-0061060.

PR 07-MAR-1997; 97US-0038621.

PR 07-MAR-1997; 97US-0040161.

PR 07-MAR-1997; 97US-0040162.

PR 07-MAR-1997; 97US-0040163.

PR 07-MAR-1997; 97US-0040333.

PR 07-MAR-1997; 97US-0040334.

PR 07-MAR-1997; 97US-0040336.

PR 07-MAR-1997; 97US-0040626.

PR 11-APR-1997; 97US-0043311.

PR 11-APR-1997; 97US-0043312.

PR 11-APR-1997; 97US-0043313.

PR 11-APR-1997; 97US-0043314.

PR 11-APR-1997; 97US-0043568.

PR 11-APR-1997; 97US-0043569.

PR 11-APR-1997; 97US-0043576.

PR 11-APR-1997; 97US-0043578.

PR 11-APR-1997; 97US-0043580.

PR 11-APR-1997; 97US-0043669.

PR 11-APR-1997; 97US-0043670.

PR 11-APR-1997; 97US-0043671.

PR 11-APR-1997; 97US-0043672.

PR 11-APR-1997; 97US-0043674.

PR 23-MAY-1997; 97US-0047492.

PR 23-MAY-1997; 97US-0047500.

PR 23-MAY-1997; 97US-0047501.

PR 23-MAY-1997; 97US-0047502.

PR 23-MAY-1997; 97US-0047503.

PR 23-MAY-1997; 97US-0047581.

PR 23-MAY-1997; 97US-0047582.

PR 23-MAY-1997; 97US-0047583.

PR 23-MAY-1997; 97US-0047584.

PR 23-MAY-1997; 97US-0047585.

PR 23-MAY-1997; 97US-0047586.

PR 23-MAY-1997; 97US-0047587.

PR 23-MAY-1997; 97US-0047588.

PR 23-MAY-1997; 97US-0047589.

PR 23-MAY-1997; 97US-0047590.

PR 23-MAY-1997; 97US-0047592.

PR 23-MAY-1997; 97US-0047593.

PR 23-MAY-1997; 97US-0047594.

PR 23-MAY-1997; 97US-0047595.

PR 23-MAY-1997; 97US-0047596.

PR 23-MAY-1997; 97US-0047597.

PR 23-MAY-1997; 97US-0047598.

PR 23-MAY-1997; 97US-0047599.

PR 23-MAY-1997; 97US-0047600.

PR 23-MAY-1997; 97US-0047601.

PR 23-MAY-1997; 97US-0047612.

PR 23-MAY-1997; 97US-0047613.

PR 23-MAY-1997; 97US-0047614.

PR 23-MAY-1997; 97US-0047615.

PR 23-MAY-1997; 97US-0047617.

PR 23-MAY-1997; 97US-0047632.

PR 23-MAY-1997; 97US-0047633.

PR 06-JUN-1997; 97US-0048964.

PR 13-JUN-1997; 97US-0048974.

PR 08-JUL-1997; 97US-0049610.

PR 16-JUL-1997; 97US-0051926.

PR 18-AUG-1997; 97US-0052874.

PR 22-AUG-1997; 97US-0055724.

PR 22-AUG-1997; 97US-0056630.

PR 22-AUG-1997; 97US-0056631.

PR 22-AUG-1997; 97US-0056632.

PR 22-AUG-1997; 97US-0056636.

PR 22-AUG-1997; 97US-0056637.

PR 22-AUG-1997; 97US-0056662.

PR 22-AUG-1997; 97US-0056664.

PR 22-AUG-1997; 97US-0056845.

PR 22-AUG-1997; 97US-0056862.

PR 22-AUG-1997; 97US-0056864.

PR 22-AUG-1997; 97US-0056872.

PR 22-AUG-1997; 97US-0056874.

PR 22-AUG-1997; 97US-0056875.

PR 22-AUG-1997; 97US-0056876.

PR 22-AUG-1997; 97US-0056877.

PR 22-AUG-1997; 97US-0056878.

PR 22-AUG-1997; 97US-0056879.

PR 22-AUG-1997; 97US-0056880.

PR 22-AUG-1997; 97US-0056881.

PR 22-AUG-1997; 97US-0056882.

PR 22-AUG-1997; 97US-0056884.

PR 22-AUG-1997; 97US-0056886.

PR 22-AUG-1997; 97US-0056887.

PR 22-AUG-1997; 97US-0056888.

PR 22-AUG-1997; 97US-0056889.

PR 22-AUG-1997; 97US-0056892.

PR 22-AUG-1997; 97US-0056893.

PR 22-AUG-1997; 97US-0056894.

PR 22-AUG-1997; 97US-0056903.

PR 22-AUG-1997; 97US-0056908.

PR 22-AUG-1997; 97US-0056909.

PR 22-AUG-1997; 97US-0056910.

PR 05-SEP-1997; 97US-0056911.

PR 05-SEP-1997; 97US-0057650.

PR 05-SEP-1997; 97US-0057669.

PR 12-SEP-1997; 97US-0057761.

XX 97US-0058785.

(HUMA-) HUMAN GENOME SCI INC.

Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;

Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;

Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;

Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;

WPI; 1998-506364/43.

P-PSDB; W74830.

New isolated human genes and the secreted polypeptide(s) they encode

- useful for diagnosis and treatment of e.g. cancers, neurological

disorders, immune diseases, inflammation or blood disorders

Claim 1: Page 333-334; 721pp; English.

This sequence represents a nucleic acid molecule designated Gene 101 from

the human cDNA clone HLQAB52 (deposited as clone ATCC 97901 and ATCC

209047) which encodes a secreted human protein. The gene can be used to

generate fusion proteins by linking to the gene to a human immunoglobulin

Fc portion (e.g. V59502) for increasing the stability of the fused

protein as compared to the human protein only.

The invention relates to 186 novel genes and their fragments (nucleic

acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which

are useful for preventing, treating or ameliorating medical conditions

e.g. by protein or gene therapy. Also, pathological conditions can be

diagnosed by determining the amount of the new polypeptides in a sample

or by determining the presence of mutations in the new polynucleotides.



CC Specific uses are described for each of the 186 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see V59511 for described  
 CC uses).

XX Sequence 1751 BP; 405 A; 477 C; 476 G; 393 T; 0 other;

Query Match 1.8%; Score 41.8; DB 19; Length 1751;  
 Best Local Similarity 61.5%; Pred. No. 0.13;  
 Matches 67; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 1292 tattaattagttattcttcttgagacagagttctcaactgtgtgcccaggctagtct 1351

Db 1750 TTTTTTTTTTTTTTTTTTTTGGACGGAGTCTCGCTCTGCGCCAGGCTGGACT 1691

Oy 1352 caaactgctgctcattgtctcaactcatcaaatgctgggtccagg 1400

Db 1690 GCGGACTGCACTGGCGGCAATCTCGGCTCACTGCAAGCTCCGCTTCCCGG 1642

RESULT 15

A28150  
 ID A28150 standard; DNA; 41684 BP.

XX A28150;

XX 29-JAN-2001 (first entry)

XX Human purH gene genomic DNA sequence.

XX Human; purH; biallelic marker; single nucleotide polymorphism; SNP;  
 KW diagnosis; prevention; treatment; prostate cancer; tumour;  
 KW chromosome 1q34-q35; ds.

XX Homo sapiens.

Key	Location/Qualifiers
5'UTR	1..2000
FT	/*tag= a
FT	2001..2096
FT	/*tag= b
FT	/*number= 1
FT	2097..2432
FT	/*tag= c
FT	/*number= 1
FT	2433..2559
FT	/*tag= d
FT	/*number= 2
FT	2560..8091
FT	/*tag= e
FT	/*number= 2
FT	8092..8168
FT	/*tag= f
FT	/*number= 3
FT	8169..9599
FT	/*tag= g
FT	/*number= 3
FT	9600..9666
FT	/*tag= h
FT	/*number= 4
FT	9667..15177
FT	/*tag= i
FT	/*number= 4
FT	15178..15266
FT	/*tag= j
FT	/*number= 5
FT	15267..15923
FT	/*tag= k
FT	/*number= 5
FT	15924..16075
FT	/*tag= l
FT	/*number= 6
FT	16076..16758
FT	/*tag= m
FT	/*number= 6
FT	16759..16915
FT	/*tag= n
FT	/*number= 7
FT	16916..22308
FT	/*tag= o
FT	/*number= 7
FT	22309..22434
FT	/*tag= p
FT	/*number= 8
FT	22435..23276
FT	/*tag= q
FT	/*number= 8
FT	23277..23384
FT	/*tag= r
FT	/*number= 9
FT	23385..24840
FT	/*tag= s
FT	/*number= 9
FT	24841..24926
FT	/*tag= t
FT	/*number= 10
FT	24927..25956
FT	/*tag= u
FT	/*number= 10
FT	25957..26046
FT	/*tag= v
FT	/*number= 11
FT	26047..28699
FT	/*tag= w
FT	/*number= 11
FT	28700..28828
FT	/*tag= x
FT	/*number= 12
FT	28829..34698
FT	/*tag= y
FT	/*number= 12
FT	34699..34791
FT	/*tag= z
FT	/*number= 13
FT	34792..36678
FT	/*tag= aa
FT	/*number= 13
FT	36679..36361
FT	/*tag= ab
FT	/*number= 14
FT	36862..39013
FT	/*tag= ac
FT	/*number= 14
FT	39014..39169
FT	/*tag= ad
FT	/*number= 15
FT	39170..39455
FT	/*tag= ae
FT	/*number= 15
FT	39456..39684
FT	/*tag= af
FT	/*number= 16
FT	replace(6491,T)
FT	/*tag= ag
FT	replace(15234,G)
FT	/*tag= ah
FT	replace(15868,T)
FT	/*tag= ai
FT	replace(16729,T)
FT	/*tag= aj
FT	replace(18311,C)
FT	/*tag= ak
FT	replace(18572,T)
FT	/*tag= al
FT	replace(22906,A)
FT	/*tag= am

```
FT variation replace(23175,T)
FT /*tag= an
FT replace(23253,G)
FT /*tag= ao
FT variation replace(26106,C)
FT /*tag= ap
FT variation replace(30464,G)
FT /*tag= aq
FT variation replace(30669,C)
FT /*tag= ar
FT variation replace(31250,A)
FT /*tag= as
FT variation replace(35148,G)
FT /*tag= at
FT variation replace(36801,G)
FT /*tag= au
FT variation replace(37286,T)
FT /*tag= av
FT variation replace(37536,C)
FT /*tag= aw
FT variation replace(39321,A)
FT /*tag= ax
FT variation replace(39689,T)
FT /*tag= ay
FT primer_bind 6137..6157
FT /*tag= az
FT primer_bind complement (6577..6597)
FT /*tag= ba
FT primer_bind 14864..14882
FT /*tag= bb
FT primer_bind complement (15292..15312)
FT /*tag= bc
FT primer_bind 15837..15855
FT /*tag= bd
FT primer_bind complement (16429..16266)
FT /*tag= be
FT primer_bind 16599..16617
FT /*tag= bf
FT primer_bind complement (17030..17049)
FT /*tag= bg
FT primer_bind 18131..18150
FT /*tag= bh
FT primer_bind complement (18592..18610)
FT /*tag= bi
FT primer_bind 22710..22727
FT /*tag= bj
FT primer_bind 23100..23118
FT /*tag= bk
FT primer_bind complement (23130..23149)
FT /*tag= bl
FT primer_bind complement (23512..23530)
FT /*tag= bm
FT primer_bind 25822..25840
FT /*tag= bn
FT primer_bind complement (26222..26241)
FT /*tag= bo
FT primer_bind 30332..30352
FT /*tag= bp
FT primer_bind complement (30754..30773)
FT /*tag= bq
FT primer_bind 30918..30935
FT /*tag= br
FT primer_bind complement (31390..31408)
FT /*tag= bs
FT primer_bind 34780..34799
FT /*tag= bt
FT primer_bind complement (35215..35233)
FT /*tag= bu
FT primer_bind 36593..36610
FT /*tag= bv
FT primer_bind complement (37017..37036)
FT /*tag= bw
FT primer_bind 37060..37080
```

\* ! \* ~

```
FT primer_bind /*tag= bx
FT complement (37541..37561)
FT /*tag= by
```

Query Match 1.8%; Score 41.8; DB 21; Length 41684;  
Best Local Similarity 66.1%; Pred. No. 0.72; Indels 2; Gaps 1;  
Matches 76; Conservative 0; Mismatches 37;

```
QY 1295 taattagttattcttcttctgagacagagctctcactgtgtgcccaggctagtctcaa 1354
      ||||| || | || | ||||| ||||| ||||| ||||| ||||| |||||
Db 4598 taattttttgtatttttggtagagacagagtttcaccatgttgccaggctgactcga 4657
      ||||| || | || | ||||| ||||| ||||| ||||| ||||| |||||

QY 1355 act--tgcgggtccatttgtctcactcactcagaatgtgggtccagggtgtgtgc 1407
      ||| || || | || | || | || | || | || | || | || | || | || |
Db 4658 actcctgacctcaagtgatcgccgccttgccctgctgggattacagggtgtgagc 4712
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: June 21, 2001, 18:00:40  
Job time: 13448 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:10:18 ; Search time 8015.43 Seconds  
(without alignments)  
4386.314 Million cell updates/sec

Title: US-09-445-201-l\_COPY\_8336\_10608  
Perfect score: 2273  
Sequence: 1 tctcgacagacgctgt.....agggcttctcttgatgcc 2273

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pi1:\*

13: gb\_pi2:\*

14: gb\_pi3:\*

15: gb\_pi4:\*

16: em\_ba1:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_higo\_hum:\*

20: em\_higo\_inv:\*

21: em\_higo\_rod:\*

22: em\_hig\_hum1:\*

23: em\_hig\_hum2:\*

24: em\_hig\_hum3:\*

25: em\_hig\_hum4:\*

26: em\_hig\_hum5:\*

27: em\_hig\_hum6:\*

28: em\_hig\_hum7:\*

29: em\_hig\_hum8:\*

30: em\_hig\_inv1:\*

31: em\_hig\_inv2:\*

32: em\_hig\_other:\*

33: em\_hig\_rod:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pi:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_v12:\*

60: gb\_v1:\*

61: gb\_v2:\*

62: gb\_v3:\*

63: gb\_v4:\*

64: gb\_v5:\*

65: gb\_v6:\*

66: gb\_v7:\*

67: gb\_v8:\*

68: gb\_v9:\*

69: gb\_v10:\*

70: gb\_v11:\*

71: gb\_v12:\*

72: gb\_v13:\*

73: gb\_v14:\*

74: gb\_v15:\*

75: gb\_v16:\*

76: gb\_v17:\*

77: gb\_v18:\*

78: gb\_v19:\*

79: gb\_v20:\*

80: gb\_v21:\*

81: gb\_v22:\*

82: gb\_v23:\*

83: gb\_v24:\*

84: gb\_v25:\*

85: gb\_v26:\*

86: gb\_v27:\*

87: gb\_v28:\*

88: gb\_v29:\*

89: gb\_v30:\*

90: gb\_v31:\*

91: gb\_v32:\*

92: gb\_v33:\*

93: gb\_v34:\*

94: gb\_v35:\*

95: gb\_v36:\*

96: gb\_v37:\*

97: gb\_v38:\*

98: em\_ba3:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	100.0	12845	9	AX002124
2	510	22.4	510	94	AF061804
3	430	18.9	430	94	AF153058
C 4	333.4	14.7	214348	66	AC021220
C 5	321	14.1	135552	65	AC019251
C 6	254.4	11.2	126330	63	AC013745
C 7	56.6	2.5	230352	64	AC016982
8	56	2.5	23108	94	AF088189

Sequence  
AF061804 Mus muscu  
AF153058 Mus muscu  
AC021220 Homo sapi  
AC019251 Homo sapi  
AC013745 Homo sapi  
AC016982 Mus muscu  
AF088189 Mus muscu

9	56	2.5	68121	94	AP001917	Mus muscu
10	56	2.5	108333	94	AP001293	Mus muscu
c 11	56	2.5	213061	67	AC023248	Mus muscu
12	56	2.5	281000	94	MMU276505	Mus muscu
13	55.8	2.5	166513	75	AC073787	Mus muscu
14	55	2.4	157996	94	AC019026	Mus muscu
15	55	2.4	196858	77	AC084409	Mus muscu
c 16	54.4	2.4	7218	91	AP000293	Mus muscu
c 17	54.4	2.4	100000	91	AP000043	Mus muscu
c 18	54.4	2.4	100000	91	AP000111	Mus muscu
c 19	54.4	2.4	100000	91	AP000187	Mus muscu
c 20	54.4	2.4	223431	75	AC074310	Mus muscu
c 21	54.4	2.4	340000	91	AP001716	Mus muscu
c 22	54.2	2.4	207358	80	AL359736	Mus muscu
c 23	54.2	2.4	207430	78	AL136963	Mus muscu
c 24	53.4	2.3	215126	76	AC083916	Mus muscu
c 25	53.2	2.3	199242	81	AL589652	Mus muscu
c 26	53.2	2.3	204755	82	AL590390	Mus muscu
c 27	53	2.3	145347	75	AC073821	Mus muscu
c 28	53	2.3	163084	82	AL590410	Mus muscu
c 29	53	2.3	195556	82	AL590140	Mus muscu
c 30	53	2.3	209173	81	AL512597	Mus muscu
c 31	53	2.3	216021	97	HUAC004787	Mus muscu
c 32	53	2.3	234257	76	AC079574	Mus muscu
c 33	53	2.3	241012	76	AC079518	Mus muscu
c 34	53	2.3	340969	81	AL512630	Mus muscu
c 35	52.8	2.3	185694	76	AC079505	Mus muscu
c 36	52.8	2.3	213674	76	AC079528	Mus muscu
c 37	52.8	2.3	231669	73	AC068067	Mus muscu
c 38	52.4	2.3	9479	68	AC023897	Mus muscu
c 39	52.4	2.3	220170	74	AC073511	Mus muscu
c 40	52.2	2.3	179992	78	AC091251	Mus muscu
c 41	52.2	2.3	202083	68	AC023833	Mus muscu
c 42	51.8	2.3	107257	94	AF289665	Mus muscu
c 43	51.8	2.3	200729	67	AC022781	Mus muscu
c 44	51.8	2.3	200849	78	AC091250	Mus muscu
c 45	51.6	2.3	190430	77	AC087873	Mus muscu

# ALIGNMENTS

RESULT 1	AX002124	12845 bp	DNA	PAT	10-MAR-2000
LOCUS	Sequence 1 from Patent WO9855638.				
DEFINITION	AX002124				
ACCESSION	AX002124.1	GI:7241839			
VERSION					
KEYWORDS	unidentified.				
SOURCE	unclassified.				
ORGANISM	1 (bases 1 to 12845)				
REFERENCE	Breier, G. and Risau, W.				
AUTHORS	REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A				
TITLE	HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES THEREOF				
JOURNAL	Patent: WO 9855638-A 1 10-DEC-1998;				
FEATURES	MAX PLANKCK GESELLSCHAFT (DE); BREIER GEORG (DE)				
Source	Location/Qualifiers				
	1..12845				
	/db_xref="taxon:32644"				
BASE COUNT	3321 a 2893 c 2863 g 3723 t	45	others		
ORIGIN					
Query Match	100.0%;	Score 2273;	DB 9;	Length 12845;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 2273;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	tctcgagacagcgctgtgttttctctcgtcttgcagcggaagataacagt	60		

Db	8336	TCTCGAGACAGAGCCGCTGTTGTTTTTCCCTTGGTCTTTTGGCGGGAAGATAACAGTG	8395
QY	61	cacaaataaataaattgttgcgatttgacataaaaggctttttatgtatagta	120
Db	8396	CACAAATTAATTANTGTTGGTATCGGATTGAACATAAAAGGCGCTTTATTGTATAGTA	8455
QY	121	gcatactacaccttgcagtcagaatgagctgtctaaagaacacccccaaacttgcgga	180
Db	8456	GCATATGTACCTCTTGCAGTCAGATGAGCTGCTAAAGAACAACACCAAACTTGCCGA	8515
QY	181	tgaataaataagaggttttaataaaaggcgatgagcattagtcactgaigtaaatctc	240
Db	8516	TGAAAATGAATGAGGTTTAAATAAGCGGATGGATGAGCATAGTCACTGATGATAAATCTC	8575
QY	241	cagttatgtataacacctcattgactgatttgattgcagacatatttggtatggggcctc	300
Db	8576	CAGTTATTGATTAACCTCATTGACTGGATTGTGTCAGACATGATTTGGTATGGGGCATC	8635
QY	301	ctttaaagatgagcatagccaaagtcctgcactctaaagaatctctatggtctgatgta	360
Db	8636	CTTTAAAGATGAGCATAGCAACGCTGCTGCACTCTAAAGAGAATCTATGGCTGTATGTTA	8695
QY	361	ttacagacacattgagaagctcttagtgcctgctgctgctgctgctgctgctgctgctgct	420
Db	8696	TTACAGACACATTGAGAAGCTCTTTAGTGGCTCTGCGCTGTAGATCAGCGGTAGAGCGCT	8755
QY	421	gaggtctgcgctgcgctctcctgcactgaagataaaagccatttactgtggtggtgcag	480
Db	8756	GAGGCTCTGCGCTCGCTTCTTGGCACTGAAGATAAAGGCCATTACTGTGGTGGTCAG	8815
QY	481	tgggcgcagtttgcagcaggttactactactactactactactactactactactactactatga	540
Db	8816	TGGGCGCAGTTTGTGACGAGTACTACTACTACTACTACTACTACTACTACTACTACTACTA	8875
QY	541	gttcacagatgagcgtaccagtgattgttgcaggttaagtgttcagtcagtcagtcagtcag	600
Db	8876	GTTTCATCAGATCAGCGGTATCCAGTGATGTTTGCAGGTTAATGTTCTCAGTCACTGTTT	8935
QY	601	gaatcactatcaaaacaaattgtttctcatttctcctgctcttctcaaaacaaagtaaga	660
Db	8936	GAATCTACTATCAACAAATTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	8995
QY	661	ttccattattgaaagcgttgcagcagcatttgcctgctgctgctgctgctgctgctgctgctgct	720
Db	8996	TTCCATTATTGAAGCGCTTGTTTAAGAGCATTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT	9055
QY	721	gactatttccattatgacaaatattatgcgattatgaatgaatgactacccagttctat	780
Db	9056	GACTTATTTCATATTGACAAATATTATGCGGATTAATTGATATGACTACCCAGTTCTAT	9115
QY	781	agctgctcagggcagacacagacatctgtgacccagtcactttaaagtcctttaaagtcctttaaag	840
Db	9116	AGCTGCTCAGGGCAGACCAAGAGCATCTGTATCAGTCACCTTAAATGCCATTTAAAA	9175
QY	841	tgataaattgtgtctaggaataaaacacactgtaaaagtttaagaatcaacggcccaaca	900
Db	9176	TGCATAAATTGTTGTTAGGAAATAAACACACTGTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAG	9235
QY	901	caagtctttaacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	960
Db	9236	CAAGTCTTTAAACAAATGCCAACTAGCTTCTGAGATTTCATTAATGATCACTTTAAATGACCAATG	9295
QY	961	ttttaaataatgctcatttaattactaaatctatagttgtaacagcaacacatglacatct	1020
Db	9296	TTTTAAAAATATGTCATTAATTAATCTAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAG	9355
QY	1021	tattaagttgggtatattcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	1080
Db	9356	TATTAAAGTTGGGTATATTTCAGGGTGGCATAGCTGTAGACTATTGACATCTGTTGTTGGT	9415
QY	1081	agccagtgagagacgtccctccctggtgttctcagaagggccacaggttcacggcattggct	1140
Db	9416	AGCCAGTGAGAACTGCCCTCCCTGGCTGTTCTCAGAGGGCCACAGTGTACAGGCATTTGGCT	9475

QY 1141 atttgcccttggtcttcttgcataactttattgacatggcctcatctcttcgttccagttccac 1200  
Db 9476 ATTTCCTTGGCTCTTTCGCTAACTATTATGACATGGCTCATCTTCGTTTCAGCTTCAC 9535  
QY 1201 ttattgccaacaacgctcaatccagctcagccttaggagtcactctgctttagtcag 1260  
Db 9536 TTATTTCGCCCAACAGCTCAATCCAGCTGAGGCTTTAGGAGTCATCTGTTCATGTCAG 9595  
QY 1261 tgcgaattagaagcctggatgctgctgctctatttaattagttattcttctctctgaga 1320  
Db 9596 TCGCAATTAGAAGCCTGGATGCTGCTGCTATTATTAATTAGTTATTCTTCTCTGAGA 9655  
QY 1321 cagagctcactgtgtgccccagcctagtcctcaaaacttgcggctccatcttctcactcat 1380  
Db 9656 CAGAGTCTCACTGTGTGCCCGAGCTAGTCTCAAACTTGGCGTCCATTCTCTCACTCAT 9715  
QY 1381 cagaatgctggcctccagctgtgtgcacacactagtgagctcgcgttttaagctaaga 1440  
Db 9716 CAGAATGCTGGCTTCCAGGTGTGTGCACCACTAGGTAGCTGCGCTTTTAAGCTAAGA 9775  
QY 1441 gctggaagatccctgatgctcttaccatggtggcctgttaccaggttagttgactgaaga 1500  
Db 9776 GCTGGAAGATCCTGATGCTCTTACCATGCTGGCATGTTACAGGTAGTTGACTGAAAA 9835  
QY 1501 ctagtctctcgtgtgttaatgacctgcagtggtgtatgtatctctcagatgctctttttgc 1560  
Db 9836 CTAGTTATCTCGCTGTGTAATGACCTGCAGTGGTATGTATCTCTCAAGATGCTTTTTCG 9895  
QY 1561 atttcaatcagtttagttaacaagttcttaagctccagcttggttatgtggcatagctcag 1620  
Db 9896 ATTTCAATCAGTTAGTGAACAGTTCTTAAGTCTCCAGCTTGTTATGGCATGAGCTCAG 9955  
QY 1621 agcttgatttaagtggtggacccctagctattgtctcattagacttacactatttta 1680  
Db 9956 AGCTTTGATTAAAGTGTGGACCCCTAGCTATGTCTATGCTCATTAAGCTTACACTATTITA 10015  
QY 1681 gtttgcctcgtgattgaatgcgatgtatgcataagccttggagatattttctctcc 1740  
Db 10016 GTTTTGTCTCGAGTTTGAATATGCTATGCTATGCTCAACTTGGGACATATTTTCTTCC 10075  
QY 1741 ccaattcctttctccatttaaatgctgtctttagagcactgctcagctctctgc 1800  
Db 10076 CCAATTCCTTTCTCCATTAAATGTGCTCTTTTGAAGCCCACTGCCCTCAGCTTCTGC 10135  
QY 1801 agctcagatcaccagaagctcgtggtcacacagcatgataaagacaatggagcgggtca 1860  
Db 10136 AGCTCAGTACCAAGAGAGCTGTGTACACAGCATGATAAAGCAATGGGACGGGTCA 10195  
QY 1861 cagtggctccctccctttcaggggtatggagcagctgtagagatgctctccagga 1920  
Db 10196 CAGTGGCTCCCGTCCCTTTCCAGGGTATGGAGACGAGCTGTAGAGAGATGCTCTCAGGGA 10255  
QY 1921 gtttccattcaatcagaatttagtcagatgtgtcattccctctgctttaccagaatgtca 1980  
Db 10256 GTTTTCATTAAATCAGCAATTTAGTTCAGATCTGTGCATCTGCTTATACAGAAATGTCA 10315  
QY 1981 gtgggctgagatcagatgaggttctcgggtttcgaatgctcccgatctcttttga 2040  
Db 10316 GTGGGCTGAGATCATCAGATGAGGTTTCATCGGGTTTCAATGTCCCGTATCCCTTTTGTGA 10375  
QY 2041 agaccttgaagtggcaacacaggaactccacctgggtggtccgtggaattgag 2100  
Db 10376 AGACCTTGAAGTTGGCAACACAGGAAACAGGAATCCACCTGGTCCCGTGAATTGCAG 10435  
QY 2101 agctgtgtgtgtgttgatgacatctgcaccattctcctgttatgacagagcttgaa 2160  
Db 10436 AGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10495  
QY 2161 ctttaactggagactggggcaagctcaatccacctttatataatgaattgctgaagagc 2220  
Db 10496 CTTTAACTGGGACTGGGGCAAGTCAATCCACCTTTATACATGAATGCTGAAGAGGC 10555

QY 2221 cttttaaaacttgagtgatgcttattatggaaggccttccctatggatcc 2273  
Db 10556 CTTTAAAACTTGGAGTGTGCATTTTATGAAAGGCTTTCCCTATTGGATCC 10608  
RESULT 2  
AF061804 AF061804 510 bp DNA ROD 23-AUG-1999  
LOCUS Mus musculus fetal liver kinase 1 (Flkl) gene, intronic enhancer.  
DEFINITION AF061804  
ACCESSION AF061804  
VERSION AF061804.1 GI:5757676  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 510)  
AUTHORS Kappel,A., Ronicke,V., Damert,A., Flamme,I., Risau,W. and Breier,G.  
TITLE Identification of vascular endothelial growth factor (VEGF) receptor-2 (Flk-1) promoter/enhancer sequences sufficient for angioblast and endothelial cell-specific transcription in transgenic mice  
JOURNAL Blood 93 (12), 4284-4292 (1999)  
MEDLINE 99290783  
PUBMED 10361126  
REFERENCE 2 (bases 1 to 510)  
AUTHORS Kappel,A., Ronicke,V., Damert,A., Flamme,I., Risau,W. and Breier,G.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-1998) Molekulare Zellbiologie, Max Planck Institut fuer physiologische und klinische Forschung, Parkstrasse 1, Bad Nauheim 61231, Germany  
FEATURES  
Location/Qualifiers  
source 1..510  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
gene 1..510  
/gene="Flkl"  
/note="fetal liver kinase 1"  
enhancer 1..510  
/gene="Flkl"  
/note="intronic"  
BASE COUNT 129 a 105 c 131 g 145 t  
ORIGIN  
Query Match 22.4%; Score 510; DB 94; Length 510;  
Best Local Similarity 100.0%; Pred. No. 1.2e-117;  
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1761 taaatgtctgtctttagaagccactgctcagcttctgcagctcagataccacaaaggaag 1820  
Db 1 TAAATGTGCTGCTTTAGAGCCACTGCTCAGCTTCTGCAGCTCAGATACCAAGGAAG 60  
QY 1821 tctgtcacagcatgataaagacaatggagcgggtcacagtggtcccgctccctttc 1880  
Db 61 TCTGTTACACAGCATGATAAAGACAATGGGACGGGTACAGTGGCTCCGTCCTTTC 120  
QY 1881 aggggtatggagcagcgttagagagatgtctccagggagttttcatatcagcaatt 1940  
Db 121 AGGGGTATGGAGCAGGCTGTAGAGATGTCTCAGGGAGTTTTTCATTAATCAGCAAT 180  
QY 1941 tagtcagatctgtcctcctctatgctttacaagaatgtcagtggtggcctgagatcaga 2000  
Db 181 TAGTCAGATCTGTGCATCTCTATGCTTTACAAGAAATGTAGTGGGCTTGAGATCAGCA 240  
QY 2001 tggaggtttcactgggtttcaatgtcccgatctccttttgaagaccttgaagtggcaagc 2060  
Db 241 TGGAGGTTCATCGGTTTCAATGTCCCGTATCTCTTTTGAAGACCTTCAAGTTGCAACG 300  
QY 2061 cagaaacacaggaactccacctggtgcgtgaattgacagagctgtgtgtgtgtgtg 2120



\* 6525 9844: contig of 3320 bp in length  
 \* 9845 9944: gap of unknown length  
 \* 9945 14101: contig of 4157 bp in length  
 \* 14102 14201: gap of unknown length  
 \* 14202 21279: contig of 7078 bp in length  
 \* 21280 21379: gap of unknown length  
 \* 21380 31346: contig of 9967 bp in length  
 \* 31347 31447: gap of unknown length  
 \* 31448 40526: contig of 9080 bp in length  
 \* 40527 40527: gap of unknown length  
 \* 40528 52862: contig of 12236 bp in length  
 \* 52863 52963: gap of unknown length  
 \* 52964 66091: contig of 13128 bp in length  
 \* 66092 66191: gap of unknown length  
 \* 66192 82659: contig of 16469 bp in length  
 \* 82660 82759: gap of unknown length  
 \* 82760 104797: contig of 22038 bp in length  
 \* 104798 104897: gap of unknown length  
 \* 104898 129319: contig of 24422 bp in length  
 \* 129320 129419: gap of unknown length  
 \* 129420 176511: contig of 47092 bp in length  
 \* 176512 176611: gap of unknown length  
 \* 176612 177666: contig of 1055 bp in length  
 \* 177667 177767: gap of unknown length  
 \* 177768 178997: contig of 1231 bp in length  
 \* 178998 179097: gap of unknown length  
 \* 179098 180242: contig of 1145 bp in length  
 \* 180243 180342: gap of unknown length  
 \* 180343 181442: contig of 1100 bp in length  
 \* 181443 181543: gap of unknown length  
 \* 181544 182642: contig of 1100 bp in length  
 \* 182643 182742: gap of unknown length  
 \* 182743 184165: contig of 1423 bp in length  
 \* 184166 184266: gap of unknown length  
 \* 184267 185112: contig of 1247 bp in length  
 \* 185113 185612: gap of unknown length  
 \* 185613 187121: contig of 1509 bp in length  
 \* 187122 187221: gap of unknown length  
 \* 187222 188524: contig of 1303 bp in length  
 \* 188525 188624: gap of unknown length  
 \* 188625 189782: contig of 1158 bp in length  
 \* 189783 189882: gap of unknown length  
 \* 189883 191312: contig of 1430 bp in length  
 \* 191313 191412: gap of unknown length  
 \* 191413 192837: contig of 1425 bp in length  
 \* 192838 192937: gap of unknown length  
 \* 192939 194117: contig of 1180 bp in length  
 \* 194118 194217: gap of unknown length  
 \* 194219 195948: contig of 1731 bp in length  
 \* 195949 196048: gap of unknown length  
 \* 196049 197744: contig of 1696 bp in length  
 \* 197745 197844: gap of unknown length  
 \* 197845 199489: contig of 1645 bp in length  
 \* 199490 199589: gap of unknown length  
 \* 199590 201304: contig of 1715 bp in length  
 \* 201305 201404: gap of unknown length  
 \* 201405 203316: contig of 1912 bp in length  
 \* 203317 203416: gap of unknown length  
 \* 203417 204490: contig of 1074 bp in length  
 \* 204491 204590: gap of unknown length  
 \* 204591 206676: contig of 2086 bp in length  
 \* 206677 206776: gap of unknown length  
 \* 206777 208803: contig of 2027 bp in length  
 \* 208804 208903: gap of unknown length  
 \* 208904 210597: contig of 1694 bp in length  
 \* 210598 210697: gap of unknown length  
 \* 210699 212097: contig of 1400 bp in length  
 \* 212098 212197: gap of unknown length  
 \* 212199 214348: contig of 2151 bp in length.  
 \* 214349 Location/Qualifiers  
 \* 1. .214348  
 \* /organism="Homo sapiens"  
 \* /db\_xref="taxon:9606"

/chromosome="4"  
 /clone="RP11-530117"  
 1. .1584  
 /note="assembly\_name:Contig100"  
 1685. .4194  
 /note="assembly\_name:Contig101"  
 4295. .6424  
 /note="assembly\_name:Contig102"  
 6525. .9844  
 /note="assembly\_name:Contig103"  
 9945. .14101  
 /note="assembly\_name:Contig104"  
 14202. .21279  
 /note="assembly\_name:Contig105"  
 21380. .31346  
 /note="assembly\_name:Contig106"  
 31447. .40526  
 /note="assembly\_name:Contig107"  
 clone\_end:77  
 vector\_side:left"  
 40627. .52862  
 /note="assembly\_name:Contig108"  
 52963. .66090  
 /note="assembly\_name:Contig109"  
 66191. .82659  
 /note="assembly\_name:Contig110"  
 82760. .104797  
 /note="assembly\_name:Contig111"  
 104898. .129319  
 /note="assembly\_name:Contig112"  
 clone\_end:SP6  
 vector\_side:right"  
 129420. .176511  
 /note="assembly\_name:Contig113"  
 176612. .177666  
 /note="assembly\_name:Contig133"  
 177767. .178997  
 /note="assembly\_name:Contig37"  
 179098. .180242  
 /note="assembly\_name:Contig56"  
 180343. .181442  
 /note="assembly\_name:Contig59"  
 181543. .182642  
 /note="assembly\_name:Contig66"  
 182743. .184165  
 /note="assembly\_name:Contig69"  
 184266. .185512  
 /note="assembly\_name:Contig75"  
 185613. .187121  
 /note="assembly\_name:Contig77"  
 187222. .188524  
 /note="assembly\_name:Contig78"  
 188625. .189782  
 /note="assembly\_name:Contig80"  
 189883. .191312  
 /note="assembly\_name:Contig85"  
 191413. .192837  
 /note="assembly\_name:Contig86"  
 192938. .194117  
 /note="assembly\_name:Contig87"  
 194218. .195948  
 /note="assembly\_name:Contig88"  
 196049. .197744  
 /note="assembly\_name:Contig89"  
 197845. .199489  
 /note="assembly\_name:Contig90"  
 199590. .201304  
 /note="assembly\_name:Contig91"  
 201405. .203316  
 /note="assembly\_name:Contig93"  
 203417. .204490

Query Match

14.7%; Score 333.4; DB 66; Length 214348;





```
* 18310 18409: gap of 100 bp in length
* 18410 19183: contig of 774 bp in length
* 19184 19283: gap of 100 bp in length
* 19284 20038: contig of 755 bp in length
* 20039 20138: gap of 100 bp in length
* 20139 20878: contig of 740 bp in length
* 20879 20978: gap of 100 bp in length
* 20979 21780: contig of 802 bp in length
* 21781 21880: gap of 100 bp in length
* 21881 22642: contig of 762 bp in length
* 22643 22742: gap of 100 bp in length
* 22743 23549: contig of 807 bp in length
* 23550 23649: gap of 100 bp in length
* 23650 24442: contig of 793 bp in length
* 24443 24542: gap of 100 bp in length
* 24543 25313: contig of 771 bp in length
* 25314 25413: gap of 100 bp in length
* 25414 26164: contig of 751 bp in length
* 26165 26264: gap of 100 bp in length
* 26265 27045: contig of 781 bp in length
* 27046 27145: gap of 100 bp in length
* 27146 27921: contig of 776 bp in length
* 27922 28021: gap of 100 bp in length
* 28022 28812: contig of 791 bp in length
* 28813 28912: gap of 100 bp in length
* 28913 29663: contig of 751 bp in length
* 29664 29763: gap of 100 bp in length
* 29764 30555: contig of 792 bp in length
* 30556 30655: gap of 100 bp in length
* 30656 31427: contig of 772 bp in length
* 31428 31527: gap of 100 bp in length
* 31528 32317: contig of 790 bp in length
* 32318 32417: gap of 100 bp in length
* 32418 33198: contig of 781 bp in length
* 33199 33298: gap of 100 bp in length
* 33299 34081: contig of 783 bp in length
* 34082 34181: gap of 100 bp in length
* 34182 34938: contig of 757 bp in length
* 34939 35038: gap of 100 bp in length
* 35039 35795: contig of 757 bp in length
* 35796 35895: gap of 100 bp in length
* 35896 36624: contig of 729 bp in length
* 36625 36724: gap of 100 bp in length
* 36725 37501: contig of 777 bp in length
* 37502 37601: gap of 100 bp in length
* 37602 38375: contig of 774 bp in length
* 38376 38475: gap of 100 bp in length
* 38476 39284: contig of 789 bp in length
* 39285 39384: gap of 100 bp in length
* 39385 40148: contig of 784 bp in length
* 40149 40248: gap of 100 bp in length
* 40249 41001: contig of 753 bp in length
* 41002 41101: gap of 100 bp in length
* 41102 41871: contig of 770 bp in length
* 41872 41971: gap of 100 bp in length
* 41972 42757: contig of 786 bp in length
* 42758 42857: gap of 100 bp in length
* 42858 43646: contig of 789 bp in length
* 43647 43746: gap of 100 bp in length
* 43747 44507: contig of 761 bp in length
* 44508 44607: gap of 100 bp in length
* 44608 45362: contig of 755 bp in length
* 45363 45462: gap of 100 bp in length
* 45463 46193: contig of 731 bp in length
* 46194 46293: gap of 100 bp in length
* 46294 47090: contig of 797 bp in length
* 47091 47190: gap of 100 bp in length
* 47191 47981: contig of 791 bp in length
* 47982 48081: gap of 100 bp in length
* 48082 48867: contig of 786 bp in length
* 48868 48967: gap of 100 bp in length
* 48968 49724: contig of 757 bp in length
* 49725 49824: gap of 100 bp in length
```

```
* 49825 50586: contig of 762 bp in length
* 50587 50686: gap of 100 bp in length
* 50687 51456: contig of 770 bp in length
* 51457 51556: gap of 100 bp in length
* 51557 52334: contig of 778 bp in length
* 52335 52434: gap of 100 bp in length
* 52435 53204: contig of 770 bp in length
* 53205 53304: gap of 100 bp in length
* 53305 54077: contig of 773 bp in length
* 54078 54177: gap of 100 bp in length
* 54178 54958: contig of 781 bp in length
* 54959 55058: gap of 100 bp in length
* 55059 55833: contig of 775 bp in length
* 55834 55933: gap of 100 bp in length
* 55934 56716: contig of 783 bp in length
* 56717 56816: gap of 100 bp in length
* 56817 57612: contig of 796 bp in length
* 57613 57712: gap of 100 bp in length
* 57713 58422: contig of 710 bp in length
* 58423 58522: gap of 100 bp in length
* 58523 59310: contig of 788 bp in length
* 59311 59410: gap of 100 bp in length
* 59411 60186: contig of 776 bp in length
* 60187 60286: gap of 100 bp in length
* 60287 61061: contig of 775 bp in length
* 61062 61161: gap of 100 bp in length
* 61162 61891: contig of 730 bp in length

Query Match 14.1%; Score 321; DB 65; Length 135552;
Best Local Similarity 77.8%; Pred. No. 5.7e-70;
Matches 444; Conservative 0; Mismatches 110; Indels 17; Gaps 4:

Qy 1717 gaacttggagagatattttctcccaattcttctccattcttccatttaagtactgtcttt 1776
Db 63618 GAATTGGGAATGTTTTCTCTCTCCCTTCTCTCCCTTCTCTCTTAAATGACAGCTCTT 63559

Qy 1777 agaagccactgcctcagctcttcgagctcagatcaccaagaagaagctggtcacagcatg 1836
Db 63558 ATGAGCCATTATTTACAGCTGTGGCAGTTTGGTTACCAAGGGGAAGGCCTAGTA-AAATTG 63500

Qy 1837 ataaaagacaatggagcggggtcacagtggtcccggtcccttccaggggttgagagaca 1896
Db 63499 ATAAAGGAAAATGAGACAAAGTCTAGATTCTCTCACTCCCTTCA-GGGTACGTAGATGA 63441

Qy 1897 gctgtagagagatgctccaggaggtttcttaataacagcaatttagtcagatctgtaca 1956
Db 63440 ACTATATAAAAATCCGCTAAGTGGGATTCGTTAATACGCAATTTAGTCAANTGTGTACA 63381

Qy 1957 tcctatgctttacaagaatgctcagtgggcc-----tgagatacatcagatg 2002
Db 63380 TCCTATGTTCTATAAGAANTGTCAGTGGGTCTCTTTCCCAAGGGAGTGAGATCATCAGATG 63321

Qy 2003 gaggttcaatcggtttcaatgctcccgatcctcttttgtaagacccttgagagtcacaca 2062
Db 63320 AAGGTTCAATTTGGTTTCAATGTCCTCGTATCTCTTTTGTAAAGACCTTTGAAGTGGCAATGCA 63261

Qy 2063 ggaataacaggaactccaccctggtgcctggaatgcagagctgtgtgtgtgtgtgtac 2122
Db 63260 GGAATAACAGGAACCTCCACCTAGCTCCATGATGCAAGTGTGTGTGTGTGTGTATGAC 63201

Qy 2123 catctgccattcttctctgttatgacagagctgttgaaactttaactggagctggggcga 2182
Db 63200 CATCTGCCCATTTCTCTCGTTATGACACAGCTTGTGAACTTTTACTGAGAAATGGTGAAAA 63141

Qy 2183 gtcaa-tcccaactttatacaaatgaattgctgaagagcccttttaaaacttggagtgac 2241
Db 63140 GTAAATTCGCCAGTTTTATACAAATGAATTTGCTGAAGAGCCCTTTTAAAGATATAGAGTATGC 63081

Qy 2242 attgttatgaagggcttctctattgcatc 2272
Db 63080 ATTGTTATGAAGGTGTTTCTTATTAGTGC 63050
```



36362 37081: contig of 720 bp in length  
37082 37181: gap of 100 bp  
37182 37898: contig of 717 bp in length  
37899 37998: gap of 100 bp  
37999 38730: contig of 732 bp in length  
38731 38930: gap of 100 bp  
38931 39526: contig of 696 bp in length  
39527 39626: gap of 100 bp  
39627 40316: contig of 690 bp in length  
40317 40416: gap of 100 bp  
40417 41136: contig of 720 bp in length  
41137 41236: gap of 100 bp  
41237 41985: contig of 749 bp in length  
41986 42085: gap of 100 bp  
42086 42815: contig of 730 bp in length  
42816 42915: gap of 100 bp  
42916 43632: contig of 717 bp in length  
43633 43732: gap of 100 bp  
43733 44445: contig of 713 bp in length  
44446 44545: gap of 100 bp  
44546 45287: contig of 742 bp in length  
45288 45387: gap of 100 bp  
45388 46129: contig of 742 bp in length  
46130 46229: gap of 100 bp  
46230 46961: contig of 732 bp in length  
46962 47061: gap of 100 bp  
47062 47765: contig of 704 bp in length  
47766 47865: gap of 100 bp  
47866 48582: contig of 717 bp in length  
48583 48682: gap of 100 bp  
48683 49429: contig of 747 bp in length  
49430 49529: gap of 100 bp  
49530 50236: contig of 707 bp in length  
50237 50336: gap of 100 bp  
50337 51062: contig of 726 bp in length  
51063 51162: gap of 100 bp  
51163 51884: contig of 722 bp in length  
51885 51984: gap of 100 bp  
51985 52702: contig of 718 bp in length  
52703 52802: gap of 100 bp  
52803 53516: contig of 714 bp in length  
53517 53616: gap of 100 bp  
53617 54329: contig of 713 bp in length  
54330 54429: gap of 100 bp  
54430 55140: contig of 711 bp in length  
55141 55240: gap of 100 bp  
55241 55930: contig of 690 bp in length  
55931 56030: gap of 100 bp  
56031 56724: contig of 694 bp in length  
56725 56824: gap of 100 bp  
56825 57628: contig of 804 bp in length  
57629 57728: gap of 100 bp  
57729 58523: contig of 795 bp in length  
58524 58623: gap of 100 bp  
58624 59387: contig of 764 bp in length  
59388 59487: gap of 100 bp

Query Match  
Best Local Similarity 11.2%; Score 254.4; DB 63; Length 126330;  
Matches 314; Conservative 0; Mismatches 144; Indels 15; Gaps 2;

QY 1815 aggaagtctgtgacacagcatgataaagacaatggagcgggtgcacagtggctccgctc 1874  
| | | | |  
Db 88902 ANNAACAAAGTTATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88843  
| | | | |  
QY 1875 ccttcagggtatgagacagctgtagagatgtctccagaggttttcattatca 1934  
| | | | |  
Db 88842 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88783  
| | | | |  
QY 1935 gcaatttagctgacatgcatctcttaccagaataatcagtgagcc----- 1987  
| | | | |  
Db 88782 GCAATTAGTCAAAATGTGACATCCCTATGTCTATAGAAAGATGTCACTGGTCC 88723  
| | | | |

QY 1988 -----tgagatcatcagatggaggttcacatcggttttcaatgtcccgatccctttgtga 2040  
| | | | |  
Db 88722 AAGGAGTGCATCATCAGATGAAGTTTCATTTGGTTTCATGTCCTCCGTATCCTTTGTA 88563  
| | | | |  
QY 2041 agaccttgaagtggcaacgcaggaacacaggaacacacccctcgtggtgcgtggaattgcag 2100  
| | | | |  
Db 88662 AGACCTTGAAGTTGGCAATGCAGGAAACAGAGAACTCCACCTAGCTCCATGAATTGCAG 88603  
| | | | |  
QY 2101 agctgttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2160  
| | | | |  
Db 88602 AACTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 88543  
| | | | |  
QY 2161 ctttaactgggactgggggcaaatgcaa-tccacactttatacaatgaattctgaagagg 2219  
| | | | |  
Db 88542 CTTTACTGAGATGGTGAAGTAAATCCACGTTTATACATGAATTCCTGAGAGAGG 88483  
| | | | |  
QY 2220 ccttttaaaacttggagtgctgcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2272  
| | | | |  
Db 88482 CCTTTAAAGTATAGATGATGCATTTGTTTATGGAAGGTGTTTCTCTATTAGTCT 88430  
| | | | |

RESULT 7  
AC016982/c  
LOCUS  
DEFINITION  
AC016982 230352 bp DNA HTG 26-FEB-2001  
Mus musculus clone RP23-416H2, WORKING DRAFT SEQUENCE, 11 unordered  
pieces.  
AC016982  
AC016982.4 GI:13123354  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
house mouse.  
SOURCE  
Mus musculus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 230352)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-416H2  
Unpublished  
2 (bases 1 to 230352)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Bohuslavsky,L., Bouckgalter,B., Brown,A., Castle,A., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,  
Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W.,  
Forrest,C., Gage,D., Gestalt,J., Gardyna,S., Grant,G., Hagos,B.,  
Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,  
Karatas,A., Klein,J., Landers,T., Lehoczy,J., Lieu,C., Locke,K.,  
Macdonald,P., Marquis,N., McEwen,P., McGurk,A., McKernan,K.,  
Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,  
Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu.X., Wyman,D., Ye.W.J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (09-DEC-1999), Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 25, 2001 this sequence version replaced gi:8072589.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5038  
Center clone name: 416\_H\_2  
----- Summary Statistics  
Sequencing vector: M13; M77815; 39% of reads  
Sequencing vector: plasmid; n/a; 61% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

\* NOTE: This is a working draft sequence. It currently  
\* consists of 11 contigs. The true order of the piece  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence,  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES

BASE COUNT	60350 a	54120 c	54747 g	60133 t	1002 others
ORIGIN					
vector_side:right					

Query Match	2.5%	Score 56.6;	DB 64;	Length 230352;
-------------	------	-------------	--------	----------------

## RESULT

KEYWORDS	
SOURCE	house mouse.
ORGANISM	Mus musculus

JOURNAL  
REFERENCE  
Unpublished  
2 (bases 1 to 23108)  
murine CD4 gene

**TITLE** Direct Submission

## FEATURES

```

1. 23100
source
/organism="Mus musculus"
/strain="B10.D2/nSnJ"
/db_xref="taxon:10090"
/chromosome="6"

```

```

JUL22(2001) 14:35/20002:143400
/gene="Cd4"

```

/gene="Cd4"

gene="Cd4"

gene  
14561. .>2310

```
misc_feature
14564
```

/note="alternate transcription start site"

```
/gene="Cd4"
```

```
misc_feature 17927. 18271
```

site 10 (DH10)"

```
/gene="Cd4"
```

ORIGIN  
BAND COORD

Query Match	2.36; Score 36; DB 94; Length 23108;
Best local similarity	58 38; Pred No 0 0025;

Malccies 98; Conservatve 0; Mismatch 70; Index 0; Gaps 0;

db 3026 TTAATTGTAATTTTACCACTCCTAAGCTATCTACTACACCTCCCTAAC 3085

DO 3026 TATTATTATTCATTATTAATTGGAGGCAGTGTCCTAACTAATGTAATGTAACAGGCTGGCCTACA 3085





```

Berlin, GERMANY
Revised by author 22-MAY-2000
3 (bases 1 to 281000)
Engemann, S.
Direct Submission
Submitted (22-NOV-2000) Engemann S., Dep. Trautner,
Max-Planck-Institut fuer Molekulare Genetik, Ihnestrasse 73, 14195
Berlin, GERMANY
4 (bases 1 to 281000)
Engemann, S.
Direct Submission
Submitted (24-JAN-2001) Engemann S., Dep. Trautner,
Max-Planck-Institut fuer Molekulare Genetik, Ihnestrasse 73, 14195
Berlin, GERMANY
On Jan 27, 2001 this sequence version replaced gi:11191799.
related accession numbers AJ300452-AJ300460
has a 1000 bp overlap with AJ271885.

FEATURES
    source
        location/Qualifiers
            1..281000
                /organism="Mus musculus domesticus"
                /strain="129 SV"
                /sub_species="domesticus"
                /db_xref="taxon:10092"
                /chromosomes="7"
                /cell_line="CJ7"
                join(<1431..1455,3536..3622,7265..7345,7478..7578,
                7906..8110,9688..9772,11978..12149,12416..12608,
                13783..13967,14347..14428,15828..15926,16308..16404,
                17790..17888,21514..21610,21996..22086,22181..22317,
                22743..22928,23228..23354,24142..24268,25136..25238,
                27169..27307)
                /gene="Obph1"
                /number=1
            1431..28364
                /gene="Obph1"
            1432..1455
                /gene="Obph1"
            /number=1
            join(1432..1455,3536..3622,7265..7345,7478..7578,
            7906..8110,9688..9772,11978..12149,12416..12608,
            13783..13967,14347..14428,15828..15926,16308..16404,
            17790..17888,21514..21610,21996..22086,22181..22317,
            22743..22928,23228..23354,24142..24268,25136..25238,
            27169..27307)
            /gene="Obph1"
            /codon_start=1
            /product="oxysterol binding protein"
            /protein_id="CAC27351.1"
            /db_xref="GI:12583596"
            /translation="MSLVPSQPGRDMESNGPSQPRDEPQTGSAKVPPLAEYRLCN
            GSDKECTPTTRVSKRDALKAKQENYRQEKRAKOLFSAITDPSVIMADSLKIRGT
            LKSWTKLWCVLKPGVLLIYKTPKVGQWGTLLHCCCLTERPSKKDGFCKLPHLDQ
            SYMAVKGKPGESVGTSTOPLPSSYLIFRAASESDGRCWLDALELALRCSLLRLSTCK
            QGRDCQGSPPADSPSLXGLPTSATIPQDLFPLNGSALENDAFSDKSERENAEDSD
            AETQHSRTKNSGSDLLSPGPGWRGTTTVEQVQELGELDTSOVETVSEENKSLM
            WLLRQLRGMGLSRVLPFTVLEPSFLGKLSDDYYHGDLLSRAAADDPVCMKLY
            LMYLSGFTFKPKGKIPYNPILGETFCRWLHPQNTSHFTYIAEQVSHHPVSAFYV
            SNRKDFCAGSITAKSKFYGNLSALDCKAKLTFLNRKEEYTLTPYAHRCGILYG
            TMTMELGGVNTCECNKQALDPLKLPFGSSANINQISGKIMSGEVLARLTGHW
            DRDVFTEKSSGTELFWTPESEVRQRLKRTVLLLEQSELESERLMQHYTRAREG
            DQHKATQESVLEAQRQARAREQSLTPWKQLFLDLPTEWRYEDLSFWDLK
            DIAQYEQDGLHTRQETMSGGTTFLGSPSRHRRKPPSSDRRLKASDQPSGHSQVTE
            SSGTSPSCPDLSDEDFVPGGSPCPRCRREVRLKMLQEAVLISIQEAQOELHRLHLS
            LSTVRGAQAPAPSLIQNPRSWFLICIFLTQCLFINYLK"
            /gene="Obph1"
            /number=2
            7265..7345
                /gene="Obph1"
            /number=3
            7478..7578
                /gene="Obph1"
            /number=4
            7906..8110
                /gene="Obph1"

mRNA
    join(<1431..1455,3536..3622,7265..7345,7478..7578,
    7906..8110,9688..9772,11978..12149,12416..12608,
    13783..13967,14347..14428,15828..15926,16308..16404,
    17790..17888,21514..21610,21996..22086,22181..22317,
    22743..22928,23228..23354,24142..24268,25136..25238,
    27169..27307)
    /gene="Obph1"
    /number=1

gene
    1431..28364
        /gene="Obph1"
    1432..1455
        /gene="Obph1"
    /number=1

exon
    1432..1455
        /gene="Obph1"
    /number=1

CDS
    join(1432..1455,3536..3622,7265..7345,7478..7578,
    7906..8110,9688..9772,11978..12149,12416..12608,
    13783..13967,14347..14428,15828..15926,16308..16404,
    17790..17888,21514..21610,21996..22086,22181..22317,
    22743..22928,23228..23354,24142..24268,25136..25238,
    27169..27307)
    /gene="Obph1"
    /codon_start=1
    /product="oxysterol binding protein"
    /protein_id="CAC27351.1"
    /db_xref="GI:12583596"
    /translation="MSLVPSQPGRDMESNGPSQPRDEPQTGSAKVPPLAEYRLCN
    GSDKECTPTTRVSKRDALKAKQENYRQEKRAKOLFSAITDPSVIMADSLKIRGT
    LKSWTKLWCVLKPGVLLIYKTPKVGQWGTLLHCCCLTERPSKKDGFCKLPHLDQ
    SYMAVKGKPGESVGTSTOPLPSSYLIFRAASESDGRCWLDALELALRCSLLRLSTCK
    QGRDCQGSPPADSPSLXGLPTSATIPQDLFPLNGSALENDAFSDKSERENAEDSD
    AETQHSRTKNSGSDLLSPGPGWRGTTTVEQVQELGELDTSOVETVSEENKSLM
    WLLRQLRGMGLSRVLPFTVLEPSFLGKLSDDYYHGDLLSRAAADDPVCMKLY
    LMYLSGFTFKPKGKIPYNPILGETFCRWLHPQNTSHFTYIAEQVSHHPVSAFYV
    SNRKDFCAGSITAKSKFYGNLSALDCKAKLTFLNRKEEYTLTPYAHRCGILYG
    TMTMELGGVNTCECNKQALDPLKLPFGSSANINQISGKIMSGEVLARLTGHW
    DRDVFTEKSSGTELFWTPESEVRQRLKRTVLLLEQSELESERLMQHYTRAREG
    DQHKATQESVLEAQRQARAREQSLTPWKQLFLDLPTEWRYEDLSFWDLK
    DIAQYEQDGLHTRQETMSGGTTFLGSPSRHRRKPPSSDRRLKASDQPSGHSQVTE
    SSGTSPSCPDLSDEDFVPGGSPCPRCRREVRLKMLQEAVLISIQEAQOELHRLHLS
    LSTVRGAQAPAPSLIQNPRSWFLICIFLTQCLFINYLK"
    /gene="Obph1"
    /number=2
    7265..7345
        /gene="Obph1"
    /number=3
    7478..7578
        /gene="Obph1"
    /number=4
    7906..8110
        /gene="Obph1"

exon
    1432..1455
        /gene="Obph1"
    /number=5
    9688..9772
        /gene="Obph1"
    /number=6
    11978..12149
        /gene="Obph1"
    /number=7
    12416..12608
        /gene="Obph1"
    /number=8
    13783..13967
        /gene="Obph1"
    /number=9
    14347..14428
        /gene="Obph1"
    /number=10
    15828..15926
        /gene="Obph1"
    /number=11
    16308..16404
        /gene="Obph1"
    /number=12
    17790..17888
        /gene="Obph1"
    /number=13
    21514..21610
        /gene="Obph1"
    /number=14
    21996..22086
        /gene="Obph1"
    /number=15
    22181..22317
        /gene="Obph1"
    /number=16
    22743..22928
        /gene="Obph1"
    /number=17
    23228..23354
        /gene="Obph1"
    /number=18
    24142..24268
        /gene="Obph1"
    /number=19
    25136..25238
        /gene="Obph1"
    /number=20
    27169..27307
        /gene="Obph1"
    /number=21
    join(30593..130688,131306..131403,132305..132456,
    140377..140517,141097..141184,142468..142590,
    145511..145624,146438..146585,146958..147085,
    147169..147274,147610..147707,147985..148174,
    148724..148792,152041..152121,154073..154154,
    155189..155255,157406..157466,157904..157988,
    159108..159408)
    /gene="Cars"
    /number=21
    30593..130688
        /gene="Cars"
    /number=5
    join(<31277..31342,35572..35662,37049..37176,46179..46291,
    47068..47141,48531..48627)
        /gene="Tnfrh1"
    join(31277..31342,35572..35662,37049..37176,46179..46291,
    47068..47141,48531..48627)
        /gene="Tnfrh1"
    35584..35662
        /gene="Tnfrh1"
    /number=1
    join(35584..35662,37049..37176,46179..46291,47068..47141,
    48531..48580)
        /gene="Tnfrh1"

exon
    35584..35662
        /gene="Tnfrh1"
    /number=1

CDS
    join(35584..35662,37049..37176,46179..46291,47068..47141,
    48531..48580)
        /gene="Tnfrh1"

```

```

/codon_start=1
/product="tumor necrosis factor receptor p60 homologue"
/protein_id="CAC27352.1"
/db_xref="GI:12583597"
/translation="MPESYFNCPCDGEYQSDYCCCKTCPSGTFVKAPCKIPHTOGQCE
KCHPCYFTCKDNLHDCLELSTCDKDONMWDASATSDRKCEQOIGLYYDPKFPESC
RPTCKPQGIPIVLOECNANTVCCSSVSNRNWLFULMLIVPCI"
exon      37049..37176
           /gene="Tnfrh1"
           /number=2
exon      46179..46291
           /gene="Tnfrh1"
           /number=3
exon      47068..47141
           /gene="Tnfrh1"
           /number=4
exon      48531..48580
           /gene="Tnfrh1"
           /number=5
mRNA      join(567525..67596,72244..72334,73711..73838,76251..76363,
           77140..77213,77826..77868,78677..78828)
           /gene="Tnfrh2"
gene       join(67525..67596,72244..72334,73711..73838,76251..76363,
           77140..77213,77826..77868,78677..78828)
           /gene="Tnfrh2"
CDS       join(72256..72334,73711..73838,76251..76363,77140..77213,
           77826..77857)

Query Match      2.5%; Score 56; DB 94; Length 281000;
Best Local Similarity 54.3%; Pred. No. 0.0028;
Matches 113; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1327 ctcaactgtggccaggtagctcaaaactgcggctcattgtctcaactcagaatt 1386
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275809 CTCACATATGAGCGAGCGAGTCTTAGCTTAGGCTCTTGCTCAGCTCTTAAT 275868

QY 1387 gctggggttcagggtgtgtgcaccacactagtgctgcgtctttaagcagctgga 1446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275869 GCTGGGATTCAGGTATATCTCCGGCAGTCTCTGCAACTGTTCAAACTGGAAATCTTTA 275928

QY 1447 agatcctgatgtccttaccatggtggcagctgttacaggttagttgactgaaactagtt 1506
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 275929 ACCACATGAGAGTGTAAAGAGGCCAGCATATGAGAGTGTAGAAACAGACATTCGAA 275988

QY 1507 atctcgtgtgtaatgacctgcagtggt 1534
      || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275989 ATGTGGCTCAATAAAGCACCTAACCGGT 276016

RESULT 13
AC073787
LOCUS      AC073787.2 GI:9256796
DEFINITION HTG; HTGS_PHASE2; HTGS_DRAFT.
ACCESSION AC073787
VERSION    AC073787.2
KEYWORDS   HTG; house mouse.
SOURCE     Mus musculus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 166513)
           DOE Joint Genome Institute.
           Sequencing of Mouse
           Unpublished
REFERENCE  2 (bases 1 to 166513)
           DOE Joint Genome Institute.
           Direct Submission
           Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
           On Jul 18, 2000 this sequence version replaced gi:8810404.
           -----Genome Center
           Center: Joint Genome Institute

```

```

Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1895746
Center clone name: RPCI-23_413L18
-----
Summary Statistics
Consensus quality: 157635 bases at least Q40
Consensus quality: 163578 bases at least Q30
Consensus quality: 164611 bases at least Q20
Estimated insert size: 183000; agarose-fp estimation
Quality coverage: 6.23 in Q20 bases; agarose-fp estimation
Quality coverage: 6.88 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2214: contig of 2214 bp in length
* 2215 15749: contig of 13435 bp in length
* 2315 15849: gap of unknown length
* 15750 31759: contig of 15910 bp in length
* 15850 31759: gap of unknown length
* 31760 31859: gap of unknown length
* 31860 47298: contig of 15439 bp in length
* 47299 47398: gap of unknown length
* 47399 50191: contig of 2793 bp in length
* 50192 50291: gap of unknown length
* 50292 53149: contig of 2858 bp in length
* 53150 53249: gap of unknown length
* 53250 64766: contig of 11517 bp in length
* 64767 64866: gap of unknown length
* 64867 67071: contig of 2204 bp in length
* 67071 67171: gap of unknown length
* 67171 87473: contig of 20303 bp in length
* 87474 87573: gap of unknown length
* 87574 97782: gap of unknown length
* 97783 115608: contig of 17826 bp in length
* 115609 127399: contig of 11691 bp in length
* 127400 127499: gap of unknown length
* 127500 128951: contig of 1452 bp in length
* 128952 129051: gap of unknown length
* 129052 149900: contig of 20848 bp in length
* 149900 151700: gap of unknown length
* 150000 151701: contig of 1701 bp in length
* 151701 166513: gap of unknown length
* 166513 166513: contig of 14713 bp in length.

FEATURES
             Location/Qualifiers
             1..166513
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="RP23-413L18"
             /clone_lib="RPCI mouse BAC library 23"
BASE COUNT  39178 a 43350 c 43117 g 39368 t 1500 others
ORIGIN

Query Match      2.5%; Score 55.8; DB 75; Length 166513;
Best Local Similarity 58.0%; Pred. No. 0.0031;
Matches 120; Conservative 0; Mismatches 82; Indels 5; Gaps 1;

QY 1285 tgccgtgattattagttattctctctctgagacagagctcactgtgtggccagg 1344
      || ||| | | | | | | | | | | | | | | | | | | | | | |
Db 82983 TGAGTGCACATTTCTCTTTCTTTTGTTCCTGAGACAGACATCCCTATGTAACCTAGG 83042

```



Oy 1345 ctatgtctcaactgcggtccattgtctcactcatcagaatgctgggcttcacaggtgtg 1404  
 Db 83043 CTGGCTCAACTAGTATCTCTGCTCAGCTCCATTCACAAATGCTGGGATTACAAATATG 83102  
 Oy 1405 tgcaccacactagtagctgcgttttaagctaaagactgaagatcctgatgtccttta 1464  
 Db 83103 TG-----CCTTAGCTGGCCCCAAGGTATATAAATCTACAGTGGCAGTGGCTAATCTTTTGA 83157

Oy 1465 ccattgtgggcattgttacaggttagtt 1491

Db 83158 AGTAGGGGTGGGTGTTGTTGTTGTT 83184

# RESULT 14

AC019026

LOCUS

DEFINITION Mus musculus chromosome 6 clone RP23-188E5 strain C57BL/6/J, 14-JUL-2000 complete sequence.

AC019026

VERSION AC019026.12 GI:8927595

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 157996)

AUTHORS Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A. and Kucherlapati, R.

TITLE Mouse High Throughput Sequencing

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 157996)

AUTHORS Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A. and Kucherlapati, R.

TITLE Direct Submission

JOURNAL

REFERENCE 3 (bases 1 to 157996)

AUTHORS Submitted (30-DEC-1999), Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

REFERENCE 4 (bases 1 to 157996)

AUTHORS Submitted (06-JUL-2000), Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

TITLE Direct Submission

JOURNAL

REFERENCE 5 (bases 1 to 157996)

AUTHORS Submitted (14-JUL-2000), Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

TITLE Direct Submission

JOURNAL

REFERENCE 6 (bases 1 to 157996)

AUTHORS Submitted (14-JUL-2000), Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

TITLE Direct Submission

JOURNAL

COMMENT On Jul 6, 2000 this sequence version replaced gi:7767645.

-----Genome Center Center:

Albert Einstein College of Medicine Center

Code: AECOM

Web site:

http://sequence.aecom.yu.edu/cgi-

bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts

Contact: jhan@sequence.aecom.yu.edu

CLONE LENGTH: This sequence represents the entire insert of this

clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550).

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST and cDNA sequences in Unigene. Genes demonstrate at least two exons flanked by consensus splice sites that maintain sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double stranded sequence for all regions. All sequence is completed to a standard of coverage with a minimum of 3 reads with no ambiguities. If the sequence coverage for a region does not meet this standard, it is indicated in the annotation as Low Coverage. Low coverage linkages are verified by PCR product size verification or verification of forward and reverse reads from clones which span the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated average error rate is less than 1 per 10,000 bases using the Consed quality parameters. Regions that do not meet this requirement are annotated as Low Quality.

## -----Summary Statistics:

Center project name: AAP

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

Assembly program: Phrap version 0.990319

Contig length: 157996

Fraction of Phrap value < 40: 0.0011076

Error rate shown by Consed: 0.13 per 10,000 bases

Number of N's in consensus: 1

## ----- Distribution of Quality < 40 Bases:

	1000	900	800	700	600	500	400	300	200	100	0
# bases	1000	900	800	700	600	500	400	300	200	100	0

FEATURES	Source	Location/Qualifiers	Phrap Value Range
repeat_region	repeat_region	1. 157996	5 10 15 20 25 30 35 40
repeat_region	repeat_region	/organism="Mus musculus"	
repeat_region	repeat_region	/db_xref="taxon:10090"	
repeat_region	repeat_region	/chromosome="6"	
repeat_region	repeat_region	/clone="RP23-188E5"	
repeat_region	repeat_region	448..743	
repeat_region	repeat_region	/rpt_family="(TG)n"	
repeat_region	repeat_region	complement(2116..2138)	
repeat_region	repeat_region	/rpt_family="ID3"	
repeat_region	repeat_region	complement(2139..2274)	
repeat_region	repeat_region	/rpt_family="B1_MM"	
repeat_region	repeat_region	complement(2275..2348)	
repeat_region	repeat_region	/rpt_family="ID3"	
repeat_region	repeat_region	complement(3344..3682)	
repeat_region	repeat_region	/rpt_family="L1_MM"	
repeat_region	repeat_region	4251..4347	
repeat_region	repeat_region	/rpt_family="CT-rich"	
repeat_region	repeat_region	4348..4459	
repeat_region	repeat_region	/rpt_family="(CATA)n"	
repeat_region	repeat_region	4631..4754	
repeat_region	repeat_region	/rpt_family="B4A"	
repeat_region	repeat_region	4807..4841	
repeat_region	repeat_region	/rpt_family="(CATA)n"	
repeat_region	repeat_region	4863..4884	

```

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: MAJP
Center clone name: RP23-355F23
----- Summary Statistics -----
Sequencing vector: ML3; L08821
Chemistry: Dye-terminator Bodipy: 72% of reads
Chemistry: Dye-terminator Big Dye: 27% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 185059 bases at least Q40
Consensus quality: 194215 bases at least Q30
Consensus quality: 199282 bases at least Q20
Estimated insert size: 186553; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)

```

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 32 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 30351: contig of 30351 bp in length  
 \* 30352 30451: gap of unknown length  
 \* 30452 48922: contig of 18471 bp in length  
 \* 48923 49022: gap of unknown length  
 \* 49023 63234: contig of 14212 bp in length  
 \* 63235 63334: gap of unknown length  
 \* 63335 74942: contig of 11608 bp in length  
 \* 74943 75042: gap of unknown length  
 \* 75043 88417: contig of 13375 bp in length  
 \* 88418 96823: contig of 8306 bp in length  
 \* 96824 96923: gap of unknown length  
 \* 96924 105059: contig of 8146 bp in length  
 \* 105070 105169: gap of unknown length  
 \* 105170 112290: contig of 7121 bp in length  
 \* 112291 112390: gap of unknown length  
 \* 112391 119958: contig of 7568 bp in length  
 \* 119959 120058: gap of unknown length  
 \* 120059 126582: contig of 6524 bp in length  
 \* 126583 131525: contig of 4843 bp in length  
 \* 131526 131626: gap of unknown length  
 \* 131626 138601: contig of 6876 bp in length  
 \* 138602 143451: contig of 4850 bp in length  
 \* 143452 143552: gap of unknown length  
 \* 143552 147862: contig of 4311 bp in length  
 \* 147863 152648: contig of 4686 bp in length  
 \* 147963 152748: gap of unknown length  
 \* 152649 157056: contig of 4308 bp in length  
 \* 152749 157057: gap of unknown length  
 \* 157057 161317: contig of 4161 bp in length  
 \* 161318 161418: gap of unknown length  
 \* 161418 165659: contig of 4242 bp in length  
 \* 165660 165759: gap of unknown length  
 \* 165760 169663: contig of 3904 bp in length  
 \* 169664 169763: gap of unknown length  
 \* 169764 173518: contig of 3755 bp in length  
 \* 173519 173618: gap of unknown length  
 \* 173619 176717: contig of 3099 bp in length  
 \* 176718 176818: gap of unknown length  
 \* 176818 179578: contig of 2761 bp in length  
 \* 179579 179679: gap of unknown length  
 \* 179679 182179: contig of 2501 bp in length  
 \* 182180 182279: gap of unknown length  
 \* 182280 184681: contig of 2402 bp in length  
 \* 184682 184781: gap of unknown length  
 \* 184782 186392: contig of 1611 bp in length  
 \* 186393 186493: gap of unknown length  
 \* 186493 188085: contig of 1592 bp in length  
 \* 188085 188185: gap of unknown length  
 \* 188185 190062: contig of 1878 bp in length  
 \* 190063 190163: gap of unknown length  
 \* 190163 191804: contig of 1641 bp in length  
 \* 191804 191904: gap of unknown length  
 \* 191904 192919: contig of 1015 bp in length  
 \* 192919 193019: gap of unknown length  
 \* 193019 194086: contig of 1068 bp in length  
 \* 194087 194187: gap of unknown length  
 \* 194187 195634: contig of 1447 bp in length  
 \* 195634 195734: gap of unknown length  
 \* 195734 196858: contig of 1125 bp in length.

FEATURES  
 source

Location/Qualifiers  
 1. .196858  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /clone="RP23-355F23"

BASE COUNT 55216 a 42662 c 41766 g 54102 t 3112 others  
 ORIGIN

Query Match 2.4%; Score 55; DB 77; Length 196858;  
 Best Local Similarity 60.3%; Pred. No. 0.0049;  
 Matches 108; Conservative 0; Mismatches 70; Indels 1; Gaps 1;  
 QY 1258 cagtgcaattagaagcgtgagcctgcctgcctgctattagttattctctctctg 1317  
 Db 176416 CACTGAGTCCAGCAACACTTGGTTGAAAATATTCCTTTAAAAAATATTTTTTTTTTGG 176475  
 QY 1318 agacagagctcactgtgtgcccaggctagctctcaaaacttgcggtccatttgtctcact 1377  
 Db 176476 AAGAAGATCTTACTATGTAGCCCGAGGCTTGCTTAAATGTGTATCATCTACCTCAGT 176535  
 QY 1378 catcagaatgctgggctccagggtgtgtgcaccacactagtagtcgcgttttaagct 1436  
 Db 176536 CTCTGGAGTGTGGGGTTACAGGTGTG-GTACCACAGCTGGCAGTTTCTCTTTAAAGAT 176593

Search completed: June 21, 2001, 18:32:40  
 Job time: 18339 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 13:27:01 ; Search time 8015.43 Seconds  
(without alignments)  
4440.347 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_8260\_10560  
Perfect score: 2301  
Sequence: 1 tgaataagatgagggttgcc.....aatgtggaagagcctttt 2301

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rol:\*  
95: gb\_rod:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2301	100.0	12845	9	AX002124
2	465	20.2	510	94	AF061804 Mus muscu
3	430	18.7	430	94	AF153058 Mus muscu
C 4	300.8	13.1	214348	66	AC021220 Homo sapi
C 5	288.4	12.5	135552	65	AC019251 Homo sapi
C 6	221.8	9.6	126330	63	AC013745 Homo sapi
C 7	56.6	2.5	230352	64	AC016982 Mus muscu
8	56	2.4	23108	94	AF088189 Mus muscu

9	56	2.4	68121	94	AP001917	Mus muscu	
10	56	2.4	108333	94	AP001293	Mus muscu	
C	11	56	2.4	213061	67	AC023248	Mus muscu
12	56	2.4	281000	94	MMU276505	Mus muscu	
13	55.8	2.4	166513	75	AC073787	Mus muscu	
14	55	2.4	157996	94	AC019026	Mus muscu	
15	55	2.4	196858	77	AC084409	Mus muscu	
C	16	54.4	2.4	7218	91	AP000293	Homo sapi
C	17	54.4	2.4	100000	91	AP000043	Homo sapi
C	18	54.4	2.4	100000	91	AP000111	Homo sapi
C	19	54.4	2.4	100000	91	AP000187	Homo sapi
C	20	54.4	2.4	223431	75	AC074310	Mus muscu
C	21	54.4	2.4	340000	91	AP001716	Homo sapi
C	22	54.2	2.4	207358	80	AL359736	Homo sapi
C	23	54.2	2.4	207430	78	AL136963	Homo sapi
24	53.4	2.3	215126	76	AC083916	Mus muscu	
25	53.2	2.3	199242	81	AL589652	Mus muscu	
26	53.2	2.3	204755	82	AL590390	Mus muscu	
C	27	53	2.3	145347	75	AC073821	Mus muscu
C	28	53	2.3	163084	82	AL590410	Homo sapi
C	29	53	2.3	195556	82	AL590140	Homo sapi
C	30	53	2.3	209173	81	AL512597	Mus muscu
C	31	53	2.3	216021	97	HUAC004787	
C	32	53	2.3	234257	76	AC079574	Mus muscu
C	33	53	2.3	241012	76	AC079518	Mus muscu
C	34	53	2.3	340969	81	AL512630	Mus muscu
C	35	52.8	2.3	185694	76	AC079505	Mus muscu
36	52.8	2.3	213674	76	AC079528	Mus muscu	
C	37	52.8	2.3	231669	73	AC068067	Mus muscu
38	52.4	2.3	9479	68	AC023897	Mus muscu	
C	39	52.4	2.3	220170	74	AC073511	Mus muscu
C	40	52.2	2.3	179992	78	AC091251	Mus muscu
41	52.2	2.3	202083	68	AC023833	Mus muscu	
42	51.8	2.3	107257	94	AF289665	Mus muscu	
43	51.8	2.3	200729	67	AC022781	Mus muscu	
C	44	51.8	2.3	200849	78	AC091250	Mus muscu
45	51.6	2.2	190430	77	AC087873	Mus muscu	

## ALIGNMENTS

```

RESULT      1
AX002124   12845 bp      DNA          PAT      10-MAR-2000
LOCUS
DEFINITION Sequence 1 from Patent WO9855638.
ACCESSION  AX002124
VERSION     AX002124.1 GI:7241839
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified
            unclassified.
REFERENCE   1 (bases 1 to 12845)
AUTHORS    Breier,G. and Risau,W.
TITLE       REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A
            HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES
            THEREOF
JOURNAL     Patent: WO 9855638-A 1 10-DEC-1998;
            MAX PLANCK GESELLSCHAFT (DE); BREIER GEORG (DE)
FEATURES
source      Location/Qualifiers
            1..12845
             /organism="unidentified"
             /db_xref="taxon:32644"
BASE COUNT  3321 a 2893 c 2863 g 3723 t 45 others
ORIGIN
Query Match      100.0%; Score 2301; DB 9; Length 12845;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2301; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Qy 1 tgaataagatggagggttcgccttttgagtgtgtaacttgcctaattggattggcgcata 60
|||||

```







*	6525	9844: contig of 3320 bp in length
*	9845	9944: gap of unknown length
*	9945	14101: contig of 4157 bp in length
*	14102	14201: gap of unknown length
*	14202	21279: contig of 7078 bp in length
*	21280	21379: gap of unknown length
*	21380	31346: contig of 9967 bp in length
*	31347	31446: gap of unknown length
*	31447	40526: contig of 9080 bp in length
*	40527	40626: gap of unknown length
*	40637	52862: contig of 12236 bp in length
*	52863	52962: gap of unknown length
*	52963	66090: contig of 13128 bp in length
*	66091	66190: gap of unknown length
*	66191	82659: contig of 16469 bp in length
*	82660	82759: gap of unknown length
*	82760	104797: contig of 22038 bp in length
*	104798	104897: gap of unknown length
*	104898	129319: contig of 24422 bp in length
*	129320	129419: gap of unknown length
*	129420	175511: contig of 47092 bp in length
*	175512	176611: gap of unknown length
*	176612	177666: contig of 1055 bp in length
*	177667	177766: gap of unknown length
*	177767	178997: contig of 1231 bp in length
*	178998	179097: gap of unknown length
*	179098	180242: contig of 1145 bp in length
*	180243	180342: gap of unknown length
*	180343	181442: contig of 1100 bp in length
*	181443	181542: gap of unknown length
*	181543	182642: contig of 1100 bp in length
*	182643	182742: gap of unknown length
*	182743	184165: contig of 1423 bp in length
*	184166	184265: gap of unknown length
*	184266	185512: contig of 1247 bp in length
*	185513	185612: gap of unknown length
*	185613	187121: contig of 1509 bp in length
*	187122	187221: gap of unknown length
*	187222	188524: contig of 1303 bp in length
*	188525	188624: gap of unknown length
*	188625	189782: contig of 1158 bp in length
*	189783	189882: gap of unknown length
*	189883	191312: contig of 1430 bp in length
*	191313	191412: gap of unknown length
*	191413	192837: contig of 1425 bp in length
*	192838	192937: gap of unknown length
*	192938	194117: contig of 1180 bp in length
*	194118	194217: gap of unknown length
*	194218	195948: contig of 1731 bp in length
*	195949	196048: gap of unknown length
*	196049	197744: contig of 1696 bp in length
*	197745	197844: gap of unknown length
*	197845	199489: contig of 1645 bp in length
*	199490	199589: gap of unknown length
*	199590	201304: contig of 1715 bp in length
*	201305	201404: gap of unknown length
*	201405	203316: contig of 1912 bp in length
*	203317	203416: gap of unknown length
*	203417	204490: contig of 1074 bp in length
*	204491	204590: gap of unknown length
*	204591	206676: contig of 2086 bp in length
*	206677	206776: gap of unknown length
*	206777	208803: contig of 2027 bp in length
*	208804	208903: gap of unknown length
*	208904	210597: contig of 1694 bp in length
*	210598	210697: gap of unknown length
*	210698	212097: contig of 1400 bp in length
*	212098	212197: gap of unknown length
*	212198	214348: contig of 2151 bp in length.
FEATURES		Location/Qualifiers
source		1. 214348

### Query Match

13.18; Score 300.8; DB 66; Length 214348;

[illegible]

Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., VO,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 5, 2000 this sequence version replaced gi:6649482.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence-submissions@genome.wi.mit.edu](mailto:sequence-submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L2005  
Center clone name: 5\_E\_20  
-----

\* NOTE: This record contains 153 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1	752:	contig of 752 bp in length
753	852:	gap of 100 bp
853	1637:	contig of 785 bp in length
1638	1737:	gap of 100 bp
1738	2506:	contig of 769 bp in length
2507	2606:	gap of 100 bp
2607	3392:	contig of 786 bp in length
3393	3492:	gap of 100 bp
3493	4261:	contig of 769 bp in length
4262	4361:	gap of 100 bp
4362	5128:	contig of 767 bp in length
5129	5228:	gap of 100 bp
5229	6000:	contig of 772 bp in length
6001	6100:	gap of 100 bp
6101	6856:	contig of 756 bp in length
6857	6956:	gap of 100 bp
6957	7734:	contig of 778 bp in length
7735	7834:	gap of 100 bp
7835	8636:	contig of 802 bp in length
8637	8736:	gap of 100 bp
8737	9526:	contig of 790 bp in length
9527	9626:	gap of 100 bp
9627	10374:	contig of 748 bp in length
10375	10474:	gap of 100 bp
10475	11244:	contig of 770 bp in length
11245	11344:	gap of 100 bp
11345	12128:	contig of 784 bp in length
12129	12228:	gap of 100 bp
12229	13009:	contig of 781 bp in length
13010	13109:	gap of 100 bp
13110	13926:	contig of 817 bp in length
13927	14026:	gap of 100 bp
14027	14758:	contig of 732 bp in length
14759	14858:	gap of 100 bp
14859	15655:	contig of 797 bp in length
15656	15755:	gap of 100 bp
15756	16543:	contig of 788 bp in length
16544	16643:	gap of 100 bp
16644	17436:	contig of 793 bp in length
17437	17536:	gap of 100 bp
17537	18309:	contig of 773 bp in length

```

* 18310 18409: gap of 100 bp in length
* 18410 19183: contig of 774 bp in length
* 19184 19283: gap of 100 bp in length
* 19284 20038: contig of 755 bp in length
* 20039 20138: gap of 100 bp in length
* 20139 20878: contig of 740 bp in length
* 20879 20978: gap of 100 bp in length
* 20979 21780: contig of 802 bp in length
* 21781 21880: gap of 100 bp in length
* 21881 22642: contig of 762 bp in length
* 22643 22742: gap of 100 bp in length
* 22743 23549: contig of 807 bp in length
* 23550 23649: gap of 100 bp in length
* 23650 24442: contig of 793 bp in length
* 24443 24542: gap of 100 bp in length
* 24543 25313: contig of 771 bp in length
* 25314 25413: gap of 100 bp in length
* 25414 26164: contig of 751 bp in length
* 26165 26264: gap of 100 bp in length
* 26265 27045: contig of 781 bp in length
* 27046 27145: gap of 100 bp in length
* 27146 27921: contig of 776 bp in length
* 27922 28021: gap of 100 bp in length
* 28022 28912: contig of 791 bp in length
* 28913 29663: contig of 751 bp in length
* 29664 29763: gap of 100 bp in length
* 29764 30555: contig of 792 bp in length
* 30556 30655: gap of 100 bp in length
* 30656 31427: contig of 772 bp in length
* 31428 31527: gap of 100 bp in length
* 31528 32317: contig of 790 bp in length
* 32318 32417: gap of 100 bp in length
* 32418 33198: contig of 781 bp in length
* 33199 33298: gap of 100 bp in length
* 33299 34081: contig of 783 bp in length
* 34082 34181: gap of 100 bp in length
* 34182 34938: contig of 757 bp in length
* 34939 35038: gap of 100 bp in length
* 35039 35795: contig of 757 bp in length
* 35796 35895: gap of 100 bp in length
* 35896 36624: contig of 729 bp in length
* 36625 36724: gap of 100 bp in length
* 36725 37501: contig of 777 bp in length
* 37502 37601: gap of 100 bp in length
* 37602 38375: contig of 774 bp in length
* 38376 38475: gap of 100 bp in length
* 38476 39264: contig of 789 bp in length
* 39265 39364: gap of 100 bp in length
* 39365 40148: contig of 784 bp in length
* 40149 40248: gap of 100 bp in length
* 40249 41001: contig of 753 bp in length
* 41002 41101: gap of 100 bp in length
* 41102 41871: contig of 770 bp in length
* 41872 41971: gap of 100 bp in length
* 41972 42757: contig of 786 bp in length
* 42758 42857: gap of 100 bp in length
* 42858 43646: contig of 789 bp in length
* 43647 43746: gap of 100 bp in length
* 43747 44507: contig of 761 bp in length
* 44508 44607: gap of 100 bp in length
* 44608 45362: contig of 755 bp in length
* 45363 45462: gap of 100 bp in length
* 45463 46193: contig of 731 bp in length
* 46194 46293: gap of 100 bp in length
* 46294 47090: contig of 797 bp in length
* 47091 47190: gap of 100 bp in length
* 47191 47981: contig of 791 bp in length
* 47982 48081: gap of 100 bp in length
* 48082 48867: contig of 786 bp in length
* 48868 48967: gap of 100 bp in length
* 48968 49724: contig of 757 bp in length
* 49725 49824: gap of 100 bp in length

```

```

* 49825 50586: contig of 762 bp in length
* 50587 50686: gap of 100 bp in length
* 50687 51456: contig of 770 bp in length
* 51457 51556: gap of 100 bp in length
* 51557 52334: contig of 778 bp in length
* 52335 52434: gap of 100 bp in length
* 52435 53204: contig of 770 bp in length
* 53205 53304: gap of 100 bp in length
* 53305 54077: contig of 773 bp in length
* 54078 54177: gap of 100 bp in length
* 54178 54958: contig of 781 bp in length
* 54959 55058: gap of 100 bp in length
* 55059 55833: contig of 775 bp in length
* 55834 55933: gap of 100 bp in length
* 55934 56716: contig of 783 bp in length
* 56717 56816: gap of 100 bp in length
* 56817 57612: contig of 796 bp in length
* 57613 57712: gap of 100 bp in length
* 57713 58422: contig of 710 bp in length
* 58423 58522: gap of 100 bp in length
* 58523 59310: contig of 788 bp in length
* 59311 59410: gap of 100 bp in length
* 59411 60186: contig of 776 bp in length
* 60187 60286: gap of 100 bp in length
* 60287 61061: contig of 775 bp in length
* 61062 61161: gap of 100 bp in length
* 61162 61891: contig of 730 bp in length

Query Match      12.5%; Score 288.4; DB 65; Length 135552;
Best Local Similarity 77.5%; Pred. No. 1.2e-60;
Matches 406; Conservative 0; Mismatches 101; Indels 17; Gaps 4;

Qy 1793 gaacttggagagatatatttcttcccccaattcttcttccatttcaatgactgtctttt 1852
Db 63618 GAATTGGGAATGTTTCTCTCTCCCTGCTTCTCTCCCTGCTTCTTAAATGACATGCTTT 63559

Qy 1853 agaagccactgcctcagctctctgcagctcagataccaaaggaagtctggtcacacagcatg 1912
Db 63558 ATGAGCCATTATTTTCAGCTGTGGCAGTTTGGTTACCGGGGAAGCGCACTAGA-AAATTG 63500

Qy 1913 ataaaagacaatggagcgggggtcacagtggtcccgctcccttccaggggtatggagacga 1972
Db 63499 ATAAAGGAAATGAGACAAGGTCTAGATTCTCTCACCTCCCTTCA-GGGTAGCTAGATGA 63441

Qy 1973 gctgtaagagatgctctccaggagtttcttaataacagcaatttagcaatctgtaca 2032
Db 63440 ACTATATANAATCCGTCTAAGTGGATTCGTTAATCAGCAATTTAGTCAAAATGTGTACA 63381

Qy 2033 tccatgctttacaagaatgtcagtgggcc-----tgagatcatcagatg 2078
Db 63380 TCCTATGTTCTATAAGAAATGTCAGTGGTCTCTTTCCCAAGGGAGTGAGATCATCAGATG 63321

Qy 2079 gaggttcaatcggtttcaatgtlcccgatcctcttttgaagacctgaagttggcaacgca 2138
Db 63320 AAGGTTCAATTTGGTTCAATGTCCCGTATCCCTTTTGTGAAGACCTTTGAAGTTGGCAATGCA 63261

Qy 2139 ggaacacaggaactccaccctgggtccgtgaattcgaagctgttgtttgtttgtac 2198
Db 63260 GGAACACAGGAACCTCCACCTAGCTCCATGAATTCGAACTGTGTGTGTGTGTATGAC 63201

Qy 2199 catctgcccattctctctgttatgacagagctgttgaaactttaactggagctggggcaca 2258
Db 63200 CATCTGCCCATTCCTCTGTTATGACACAGCTGTGTGAACCTTTTACTGAGAAATGGTGAAA 63141

Qy 2259 glcaa-tcccacotttatacaatgaattgctgaagagcctttt 2301
Db 63140 GTAAATTCACAGTCTTATACAAATGAATGCTCTGAAGAGGCTTTT 63097

```

RESULT 6  
 AC013745/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome 4 clone RP11-92P18 map 4, LOW-PASS SEQUENCE

AC013745 126330 bp DNA HTG 13-JUL-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 4 clone RP11-92P18 map 4, LOW-PASS SEQUENCE

```

ACCESSION  AC013745
VERSION     AC013745.3  GI:9114527
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 126330)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 4, clone RP11-92P18
JOURNAL     unpublished
REFERENCE   2 (bases 1 to 126330)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckghalter,B.,
            Brown,A., Castelle,A., Colangelo,M., Collins,S., Collumore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (15-NOV-1999), Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:6910837.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L2025
            Center clone name: 92_P_18
            -----
            * NOTE: This record contains 147 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1
            * 723 822: contig of 722 bp in length
            * 823 822: gap of 100 bp
            * 823 1542: contig of 720 bp in length
            * 1543 1642: gap of 100 bp
            * 1643 2371: contig of 729 bp in length
            * 2372 2471: gap of 100 bp
            * 2472 3212: contig of 741 bp in length
            * 3213 3312: gap of 100 bp
            * 3313 4049: contig of 737 bp in length
            * 4050 4149: gap of 100 bp
            * 4150 4840: contig of 691 bp in length
            * 4841 4940: gap of 100 bp
            * 4941 5676: contig of 736 bp in length
            * 5677 5776: gap of 100 bp
            * 5777 6521: contig of 745 bp in length
            * 6522 6621: gap of 100 bp
            * 6622 7344: contig of 723 bp in length
            * 7345 7444: gap of 100 bp
            * 7445 8161: contig of 717 bp in length
            *
            * 8162 8261: gap of 100 bp
            * 8262 9014: contig of 753 bp in length
            * 9015 9114: gap of 100 bp
            * 9115 9827: contig of 713 bp in length
            * 9828 9927: gap of 100 bp
            * 9928 10650: contig of 723 bp in length
            * 10651 10750: gap of 100 bp
            * 10751 11479: contig of 729 bp in length
            * 11480 11579: gap of 100 bp
            * 11580 12318: contig of 739 bp in length
            * 12319 12418: gap of 100 bp
            * 12419 13137: contig of 719 bp in length
            * 13138 13237: gap of 100 bp
            * 13238 13299: contig of 722 bp in length
            * 13960 14059: gap of 100 bp
            * 14060 14785: contig of 726 bp in length
            * 14786 14885: gap of 100 bp
            * 14886 15580: contig of 695 bp in length
            * 15581 15680: gap of 100 bp
            * 15681 16420: contig of 740 bp in length
            * 16421 16520: gap of 100 bp
            * 16521 17233: contig of 713 bp in length
            * 17234 17333: gap of 100 bp
            * 17334 18057: contig of 724 bp in length
            * 18058 18157: gap of 100 bp
            * 18158 18874: contig of 717 bp in length
            * 18875 18974: gap of 100 bp
            * 18975 19716: contig of 742 bp in length
            * 19717 19816: gap of 100 bp
            * 19817 20557: contig of 741 bp in length
            * 20558 20657: gap of 100 bp
            * 20658 21379: contig of 722 bp in length
            * 21380 21479: gap of 100 bp
            * 21480 22193: contig of 714 bp in length
            * 22194 22293: gap of 100 bp
            * 22294 23013: contig of 720 bp in length
            * 23014 23113: gap of 100 bp
            * 23114 23851: contig of 738 bp in length
            * 23852 23951: gap of 100 bp
            * 23952 24681: contig of 730 bp in length
            * 24682 24781: gap of 100 bp
            * 24782 25505: contig of 724 bp in length
            * 25506 25605: gap of 100 bp
            * 25606 26335: contig of 730 bp in length
            * 26336 26435: gap of 100 bp
            * 26436 27158: contig of 723 bp in length
            * 27159 27258: gap of 100 bp
            * 27259 27960: contig of 702 bp in length
            * 27961 28060: gap of 100 bp
            * 28061 28775: contig of 715 bp in length
            * 28776 28875: gap of 100 bp
            * 28876 29586: contig of 711 bp in length
            * 29587 29686: gap of 100 bp
            * 29687 30394: contig of 708 bp in length
            * 30395 30494: gap of 100 bp
            * 30495 31246: contig of 752 bp in length
            * 31247 31346: gap of 100 bp
            * 31347 32090: contig of 744 bp in length
            * 32091 32190: gap of 100 bp
            * 32191 32894: contig of 704 bp in length
            * 32895 32994: gap of 100 bp
            * 32995 33741: contig of 747 bp in length
            * 33742 33841: gap of 100 bp
            * 33842 34602: contig of 761 bp in length
            * 34603 34702: gap of 100 bp
            * 34703 35440: contig of 738 bp in length
            * 35441 35540: gap of 100 bp
            * 35541 36261: contig of 721 bp in length
            * 36262 36361: gap of 100 bp
            * 36362 37081: contig of 720 bp in length
            * 37082 37181: gap of 100 bp
            * 37182 37898: contig of 717 bp in length
            * 37899 37998: gap of 100 bp

```



Insert size: 229352; sum-of-contigs  
Quality coverage: 9.5 in Q20 bases; agarose-fp  
Quality coverage: 8.8 in Q20 ba.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 4159: contig of 4159 bp in length  
\* 4160 4259: gap of 100 bp  
\* 4260 6386: contig of 2127 bp in length  
\* 6387 6486: gap of 100 bp  
\* 6487 59418: contig of 52932 bp in length  
\* 59419 59518: gap of 100 bp  
\* 59519 59698: contig of 10180 bp in length  
\* 59699 69798: gap of 100 bp  
\* 69799 83694: contig of 13896 bp in length  
\* 83695 83794: gap of 100 bp  
\* 83795 101615: contig of 17821 bp in length  
\* 101616 101715: gap of 100 bp  
\* 101716 127243: contig of 25528 bp in length  
\* 127244 127343: gap of 100 bp  
\* 127344 158706: contig of 31363 bp in length  
\* 158707 158806: gap of 100 bp  
\* 158807 191011: contig of 32205 bp in length  
\* 191012 191111: gap of 100 bp  
\* 191112 222984: contig of 31873 bp in length  
\* 222985 223084: gap of 100 bp  
\* 223085 230352: contig of 7268 bp in length.

# FEATURES

source  
1. .230352  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. .4159  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
4260. .6386  
/note="assembly\_fragment"  
6487. .59418  
/note="assembly\_fragment"  
59519. .69698  
/note="assembly\_fragment"  
69799. .83694  
/note="assembly\_fragment"  
83795. .101615  
/note="assembly\_fragment"  
101716. .127243  
/note="assembly\_fragment"  
127344. .158706  
/note="assembly\_fragment"  
158807. .191011  
/note="assembly\_fragment"  
191112. .222984  
/note="assembly\_fragment"  
223085. .230352  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:right"  
BASE COUNT 60350 a 54120 c 54747 g 60133 t 1002 others  
ORIGIN

Query Match 2.5%; Score 56.6; DB 64; Length 230352;  
Best Local Similarity 71.8%; Pred. No. 0.0025;  
Matches 74; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 1381 ttcttcttctgagacagagtctcactgtgtgcccagctagtctcaaaacttgcggtc 1440

Db 110701 TTTTCTCTTTTGTGAGACAGTCTCAGTGTGTAACCCAGGCTGCTCTCAAGCTTGTGGCC 110642  
QY 1441 catttgtccacatcacaagtctggtctccagatgtgtgc 1483  
Db 110641 CCTCTGTTTTCGTCCTCCCAAGTCTGGGATTACAAGTGTGTAC 110599  
RESULT 8  
AF088189 AF088189 23108 bp DNA ROD 28-FEB-1999  
LOCUS Mus musculus CD4 antigen (Cd4) gene, partial sequence.  
DEFINITION AF088189  
ACCESSION AF088189  
VERSION AF088189.1 GI:4322490  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 23108)  
AUTHORS Hansen,R.K., McCreedy,P.M. and Sands,J.F.  
TITLE Sequence Analysis of the 5' Flanking Region and First Intron of the  
MURINE CD4 Gene  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 23108)  
AUTHORS Hansen,R.K., McCreedy,P.M. and Sands,J.F.  
TITLE Direct Submission  
JOURNAL Submitted (31-AUG-1998) Biochemistry, Loma Linda, MT219, Loma  
Linda, CA 92354, USA  
FEATURES  
source  
1. .23108  
/organism="Mus musculus"  
/strain="B10.D2/nSnJ"  
/db\_xref="taxon:10090"  
/chromosome="6"  
join(14561. .14699,23082. .>23108)  
/gene="Cd4"  
/product="CD4 antigen"  
join(14561. .14699,23082. .>23108)  
/gene="Cd4"  
14561. .14699  
/gene="Cd4"  
/number=1  
14561. .>23108  
/gene="Cd4"  
14564  
/gene="Cd4"  
/note="alternate transcription start site"  
14569  
/gene="Cd4"  
/note="alternate transcription start site"  
17927. .18271  
/gene="Cd4"  
/note="first intron silencer; DNase I hypersensitivity  
site 10 (DH10)"  
23082. .>23108  
/gene="Cd4"  
/number=2  
BASE COUNT 6086 a 5263 c 5514 g 6245 t  
ORIGIN

Query Match 2.4%; Score 56; DB 94; Length 23108;  
Best Local Similarity 58.3%; Pred. No. 0.0034;  
Matches 98; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
QY 1371 taattagttattctctctctgagacagatctactgtgtgcccaggtagtctcaa 1430  
Db 3026 TATTATTATTATTATTATTATTGGAGCAGTGTCTAACTATGATGATGAGCTGCCTACA 3085  
QY 1431 acttgcggtccattgtctcactcacaagaatgctgggtccaggtgtgtgtgcaccacac 1490  
Db 3086 ACTCTCTCTTCCTGCTAGTCTGACCCCTCCAGGACTGGGGTTACAGGTGTGAGGGCACAAC 3145









```
/translation="MPESYFNCPDGEVQSNVCCKTCPSGTFFVKAPCKIPHTGQOCE
KCHPFTGTGNDGLHDCGLSCDCKQNMWADCSATSDRKCCEQIGLYIDPKFPESC
RPTCKPQGIPIVLQECNSTATVTCSSVSNNRNWLFLLMLIVFCI"
exon      37049..37176
/gene="Tnfrh1"
/number=2
exon      46179..46291
/gene="Tnfrh1"
/number=3
exon      47068..47141
/gene="Tnfrh1"
/number=4
exon      48531..48580
/gene="Tnfrh1"
/number=5
mRNA      join(<67525..67596,72244..72334,73711..73838,76251..76363,
77140..77213,77826..77868,78677..78828)
/gene="Tnfrh2"
gene       join(67525..67596,72244..72334,73711..73838,76251..76363,
77140..77213,77826..77868,78677..78828)
/gene="Tnfrh2"
CDS        join(72256..72334,73711..73838,76251..76363,77140..77213,
77826..77857)

Query Match      2.4%; Score 56; DB 94; Length 281000;
Best Local Similarity 54.3%; Pred. No. 0.0035;
Matches 113; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1403 ctcaactgtgtgcccagctagctcacaactgcggtcattgtctctcactcatcagaat 1462
||||| ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275809 CTCACATGATGAGCCAGGCGACTCTAGCTTAAGTCTCTTGCTCAGCCTCTTAAT 275868

QY 1463 gctgggttcagtggtgtgcaccacactaggtagctgcgtttttaagctgaagctgga 1522
||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275869 GCTGGGATTCAGGTTGATATCCCGGCAGTCTCTGCAACTGTTCAACATCGGAATCTTTA 275928

QY 1523 agatccctgatctcttaccatgttggtgacgtgttacaggttagtgactgaaactagtt 1582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275929 ACCAATAGAGATGTTTAAGAGCCCAATATGAGATGAGTGTGTAGAAACAGACATTCGAA 275988

QY 1583 atctcgtgtgtaataacacgtcagtggt 1610
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275989 ATGTGGTCAATAAAGCACCTAACCGT 276016

RESULT 13
AC073787
LOCUS      AC073787.2 GI:9256796
DEFINITION HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 166513)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
REFERENCE   2 (bases 1 to 166513)
AUTHORS    DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810404.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
```

```
Center Project Name: 1895746
Center clone name: RPCI-23_413L18
-----
Summary Statistics
Consensus quality: 157635 bases at least Q40
Consensus quality: 163578 bases at least Q30
Consensus quality: 164611 bases at least Q20
Estimated insert size: 183000; agarose-fp estimation
Quality coverage: 6.23 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.88 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 2215 2214: contig of 2214 bp in length
2315 15749: contig of 13435 bp in length
15850 15849: gap of unknown length
31750 31759: contig of 15910 bp in length
31760 31859: gap of unknown length
31860 47299: contig of 15439 bp in length
47300 47399: gap of unknown length
50192 50191: contig of 2793 bp in length
50292 50291: gap of unknown length
53150 53149: contig of 2858 bp in length
53250 53249: gap of unknown length
64767 64766: contig of 11517 bp in length
64867 64866: gap of unknown length
67071 67070: contig of 2204 bp in length
67171 67170: gap of unknown length
84774 84773: contig of 20303 bp in length
87574 87573: gap of unknown length
97683 97682: contig of 10109 bp in length
97783 97782: gap of unknown length
115609 115608: contig of 17826 bp in length
115709 115708: gap of unknown length
127400 127399: contig of 11691 bp in length
127500 127501: gap of unknown length
128952 128951: contig of 1452 bp in length
129052 129051: gap of unknown length
149899 149898: contig of 20848 bp in length
150000 149999: gap of unknown length
151701 151700: contig of 1701 bp in length
151801 151800: gap of unknown length
166513 166512: contig of 14713 bp in length.

FEATURES
            Location/Qualifiers
             1..166513
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-413L18"
BASE COUNT 39178 a 43350 c 43117 g 39368 t 1500 others
ORIGIN
```

```
Query Match      2.4%; Score 55.8; DB 75; Length 166513;
Best Local Similarity 58.0%; Pred. No. 0.0039;
Matches 120; Conservative 0; Mismatches 82; Indels 5; Gaps 1;

QY 1361 tgctgtctataatagttattctctctctgagacagactctcactgtgtggccagg 1420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82983 TGAGTGCACATTTCTTCTTTTGTTCCTGAGACAGACTCCCTATGTACCTAGG 83042

QY 1421 ctagtctcaaaccttggtccattgtctcactcatcagaatgctgggtccagggtgtg 1480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83043 CTGGCTCCCAACTAGTGTCTCTGCTCAGCCTTCCAAATGCTGGATTACAAATATG 83102
```

Qy 1481 tgcaccacactagtagctgcgttttaagctaagagctggaagactcgtatgctcttta 1540  
 Db 83103 TG-----CCTTAGCTGGCCCAAGGATATAATCTACAGTGGCAGCTGCTAATCTTTTGA 83157

Qy 1541 ccattggggcattatcaggttaatt 1567  
 Db 83158 AATAGGGGGTGGGTGCTGTTGTTGTT 83184

RESULT 14  
 AC019026  
 LOCUS  
 DEFINITION Mus musculus chromosome 6 clone RP23-188E5 strain C57BL6/J,  
 complete sequence.  
 AC019026  
 VERSION  
 AC019026.12 GI:8927595  
 KEYWORDS HTG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 157996)  
 AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
 Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and  
 Kucheralapati,R.  
 TITLE Mouse High Throughput Sequencing  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 157996)  
 AUTHORS Montgomery,K.T., Grills,G., Han,J., Leé,E., Long,J., Pomerantz,R.,  
 Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and  
 Kucheralapati,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-DEC-1999) Department of Molecular Genetics, Albert  
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
 Bronx, NY 10461, USA  
 REFERENCE 3 (bases 1 to 157996)  
 AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
 Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and  
 Kucheralapati,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUL-2000) Department of Molecular Genetics, Albert  
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
 Bronx, NY 10461, USA  
 REFERENCE 4 (bases 1 to 157996)  
 AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
 Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A. and  
 Kucheralapati,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JUL-2000) Department of Molecular Genetics, Albert  
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
 Bronx, NY 10461, USA  
 COMMENT On Jul 6, 2000 this sequence version replaced gi:7767645.  
 -----Genome Center Center:  
 Albert Einstein College of Medicine Center  
 Code: AECOM  
 Web site:  
 http://sequence.aecom.yu.edu/cgi-  
 bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.htm  
 Contact: jhan@sequence.aecom.yu.edu

CLONE LENGTH: This sequence represents the entire insert of this  
 clone unless otherwise noted. If there are overlapping clones, the  
 overlaps are noted in the beginning and end of the Features  
 listing.

ANNOTATION OF FEATURES:  
 STSS are identified using ePCR (Genome Res. 7:541-550).  
 Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.  
 Genes and Regions of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST  
 and cDNA sequences in Unigene. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintain sequence continuity

across the splice junctions. Sequences that are not identical  
 matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double  
 stranded sequence for all regions. All sequence is completed to a  
 standard of coverage with a minimum of 3 reads with no ambiguities.  
 If the sequence coverage for a region does not meet this standard,  
 it is indicated in the annotation as Low Coverage. Low coverage  
 linkages are verified by PCR product size verification or  
 verification of forward and reverse reads from clones which span  
 the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated average error rate is less than 1 per 10,000  
 bases using the Consed quality parameters. Regions that do not  
 meet this requirement are annotated as Low Quality.

# -----Summary Statistics:

Center project name: AAP  
 Sequencing vector: pUC18; L08752  
 Chemistry: Dye-terminator Big Dye; 100%  
 Assembly program: Phrap version 0.990319  
 Contig length: 157996  
 Fraction of Phrap value < 40: 0.0011076  
 Error rate shown by Consed: 0.13 per 10,000 bases  
 Number of N's in consensus: 1

## ----- Distribution of Quality < 40 Bases:

	5	10	15	20	25	30	35	40
#	1000	900	800	700	600	500	400	300
bases	1000	900	800	700	600	500	400	300
	0	*	*	*	*	*	*	*

Phrap Value Range

FEATURES	Source	Location/Qualifiers
repeat_region	1..157996	/Organism="Mus musculus"
repeat_region	1..157996	/db_xref="taxon:10090"
repeat_region	1..157996	/chromosome="6"
repeat_region	1..157996	/Clones="RP23-188E5"
repeat_region	1..157996	448..743
repeat_region	1..157996	/rpt_family="(TG)n"
repeat_region	1..157996	complement(2116..2138)
repeat_region	1..157996	/rpt_family="ID3"
repeat_region	1..157996	complement(2139..2274)
repeat_region	1..157996	/rpt_family="B1_MM"
repeat_region	1..157996	complement(2275..2348)
repeat_region	1..157996	/rpt_family="ID3"
repeat_region	1..157996	complement(3344..3682)
repeat_region	1..157996	/rpt_family="I1_MM"
repeat_region	1..157996	4251..4347
repeat_region	1..157996	/rpt_family="CT-rich"
repeat_region	1..157996	4348..4459
repeat_region	1..157996	/rpt_family="(CATA)n"
repeat_region	1..157996	4631..4754
repeat_region	1..157996	/rpt_family="B4A"
repeat_region	1..157996	4807..4841
repeat_region	1..157996	/rpt_family="(CATA)n"
repeat_region	1..157996	4863..4884
repeat_region	1..157996	/rpt_family="AT-rich"
repeat_region	1..157996	5089..5164
repeat_region	1..157996	/rpt_family="AT-rich"
repeat_region	1..157996	5220..5293







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 14:16:32 ; Search time 346.98 Seconds  
(without alignments)  
3871.349 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_8260\_10560  
Perfect score: 2301  
Sequence: 1 tgaataagatgaggttgc.....aattgctgaagagcctttt 2301

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0401.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2301	100.0	12845	20 V84274	Mouse endothelial
2	55.4	2.4	5513	21 C63529	Slit protein codin
3	46	2.0	4064	21 A07587	Mouse piwi gene, d
4	45.6	2.0	51259	18 X83007	Partial mouse WRN
5	45	2.0	2362	21 Z50462	Human tissue trans
6	44	1.9	1110	19 V29343	Calcium ion channe
7	43.6	1.9	810	20 Z16495	Human gene express
8	43.2	1.9	11901	20 X02998	Human IL-1ra BAC c
9	42.6	1.9	8966	20 Z09581	Human Apo A1 genom
10	42.2	1.8	14507	18 T73568	Expression augment
11	42.2	1.8	14507	21 Z88869	Chinese hamster 2A

c 12	42	1.8	687	20	X37433	Human secreted pro
c 13	41.8	1.8	833	21	C98178	Human colon cancer
c 14	41.8	1.8	1751	19	V59610	Human secreted pro
c 15	41.8	1.8	41684	21	A28150	Human purH gene ge
c 16	41.6	1.8	10120	20	Z23683	Human DKC1 DNA fra
c 17	41.6	1.8	119950	20	X90201	Human yes1 gene.
c 18	41.4	1.8	32351	21	F21307	Human low adenosin
c 19	41.4	1.8	32351	21	A35185	Human adenosine re
c 20	41.4	1.8	40298	21	F21311	Human low adenosin
c 21	41.4	1.8	40298	21	A35189	Human adenosine re
c 22	41.2	1.8	1502	21	C59687	Human secreted pro
c 23	41	1.8	84607	20	X90847	Human PACAP genom
c 24	41	1.8	1082138	21	F22305	Arabidopsis thalia
c 25	40.8	1.8	837	20	X37525	Human secreted pro
c 26	40.8	1.8	1863	20	X85010	Human secreted pro
c 27	40.8	1.8	4116	20	X02983	Human IL-1ra BAC c
c 28	40.8	1.8	4736	20	X23315	Mouse I-alpha-OHas
c 29	40.6	1.8	4034	20	X35011	Human glucose 6-ph
c 30	40.6	1.8	6511	16	Q95493	Human Cdn-2 DNA.
c 31	40.4	1.8	209	16	T26063	Human gene signatu
c 32	40.4	1.8	267	21	C29151	Human secreted pro
c 33	40.4	1.8	267	21	C29295	Human secreted pro
c 34	40.4	1.8	1838	20	X27332	Human secreted pro
c 35	40.4	1.8	4696	20	X03041	Human IL-1ra BAC c
c 36	40.2	1.7	1035	21	C60001	Human secreted pro
c 37	40.2	1.7	6901	18	T78853	Human lecithin-cho
c 38	40.2	1.7	49999	20	Z23891	Murine LOBO genomi
c 39	40.2	1.7	49999	20	Z23896	Murine LOBO genomi
c 40	40.2	1.7	138169	21	A34791	Human adenosine re
c 41	40.2	1.7	141589	21	F20913	Human ELAM-1 polyn
c 42	40.2	1.7	141589	21	F21127	Human low adenosin
c 43	40.2	1.7	141589	21	F21152	Human low adenosin
c 44	40.2	1.7	141589	21	A35005	Human adenosine re
c 45	40.2	1.7	141589	21	A35030	Human adenosine re

#### ALIGNMENTS

```

RESULT 1
V84274
ID V84274 standard; DNA; 12845 BP.
XX
AC V84274;
XX
DT 12-APR-1999 (first entry)
XX
DE Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.
XX
KW Flk-1; endothelial growth factor receptor-2; VEGF; transcription;
KW endothelium; enhancer; vascular disease; angiogenesis; cancer;
KW diabetic retinopathy; rheumatoid arthritis; wound healing;
KW vulnery; atherosclerosis; tumour; neuronal disorder; therapy;
KW diagnosis; mouse; ss.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Promoter 6036..6959
XX Exon /*tag= a
XX 6661..7026
XX /*tag= b
XX /*number= 1
XX /*codon_start= 6960
XX Intron 7027..10632
XX /*tag= c
XX /*number= 1
XX /*note= "DNA sequences comprising nucleotides
XX 8260-10560, 8336-10608 and/or 10094-10608
XX 10137..10155
XX /*tag= d
XX /*note= "GATA/PEA3 consensus binding site"
XX
XX protein_bind
XX

```





Db 8980 ctcaacaaaagtaagattccattattgaagcgttgtttaagagcatttttaactgttcg 9039  
Qy 781 ctatgttaggacagtgacttatttcatattgacaaatattatgccgattaatgaatat 840  
Db 9040 ctatgttaggacagtgacttatttcatattgacaaatattatgccgattaatgaatat 9099  
Qy 841 gactaccagttctatagctgtctcagggcagacaaagagacatctgtgatccagtcactt 900  
Db 9100 gactaccagttctatagctgtctcagggcagacaaagagacatctgtgatccagtcactt 9159  
Qy 901 taaatgccatttaaaatgcaataatttcttggtctaggaataaaacacacactgtaaaagttag 960  
Db 9160 taaatgccatttaaaatgcaataatttcttggtctaggaataaaacacacactgtaaaagttag 9219  
Qy 961 aatcacggcccaacacaaagctcttaacaatgccaaactagcttctgagattcattaatgt 1020  
Db 9220 aatcacggcccaacacaaagctcttaacaatgccaaactagcttctgagattcattaatgt 9279  
Qy 1021 catttaattaccaatgttttaaaatattgtcttaataattacttaatactatagttgaacag 1080  
Db 9280 catttaattaccaatgttttaaaatattgtcttaataattacttaatactatagttgaacag 9339  
Qy 1081 caacacatgtacattctattaaatttggttatattcaggggtggcagtagctgtagactattg 1140  
Db 9340 caacacatgtacattctattaaatttggttatattcaggggtggcagtagctgtagactattg 9399  
Qy 1141 cacatctgtgttgtagccagtgagaaactgctctctggtcttccagaaagccacag 1200  
Db 9400 cacatctgtgttgtagccagtgagaaactgctctctggtcttccagaaagccacag 9459  
Qy 1201 tgtcacggcattggctatttgcttgccttcttcttcttaatacttattgacatggcctcat 1260  
Db 9460 tgtcacggcattggctatttgcttgccttcttcttcttcttaatacttattgacatggcctcat 9519  
Qy 1261 ctctgttcaagttcacttatttgcacaaacgctcaatgctcagctgagccttaaggagtc 1320  
Db 9520 ctctgttcaagttcacttatttgcacaaacgctcaatgctcagctgagccttaaggagtc 9579  
Qy 1321 atctgttcttagtcagtcgaatagaaagcctgagtgcttgcctgctgtatttaattagttta 1380  
Db 9580 atctgttcttagtcagtcgaatagaaagcctgagtgcttgcctgctgtatttaattagttta 9639  
Qy 1381 ttcttcttcttgagacagagtcactgtgtgtgcccagggctagcttcaaaacttgcggctc 1440  
Db 9640 ttcttcttcttgagacagagtcactgtgtgtgcccagggctagcttcaaaacttgcggctc 9699  
Qy 1441 catttctcactcatcagaatgctggcttccaggtgtgtgcccacacactagtagctc 1500  
Db 9700 catttctcactcatcagaatgctggcttccaggtgtgtgcccacacactagtagctc 9759  
Qy 1501 gcgttttaagcctgaagagctgaagatcctgatgtcccttaccatgggtggcagatgtacag 1560  
Db 9760 gcgttttaagcctgaagagatcctgatgtcccttaccatgggtggcagatgtacag 9819  
Qy 1561 gttatgtgactgaaactagttatctcgtgtgttaataagctgcagtggtatgtatctct 1620  
Db 9820 gttatgtgactgaaactagttatctcgtgtgttaataagctgcagtggtatgtatctct 9879  
Qy 1621 caaagatcttttttgcatttcaactcagttaggttaacaaagtctttaaagctccagcttgg 1680  
Db 9880 caaagatcttttttgcatttcaactcagttaggttaacaaagtctttaaagctccagcttgg 9939  
Qy 1681 attggcagtagctcagagcttgaatgaatgagttgggaccccttagctattgtcattag 1740  
Db 9940 attggcagtagctcagagcttgaatgaatgagttgggaccccttagctattgtcattag 9999  
Qy 1741 acttacactatttttagtttgcctctagtttatgaatgcatgtatgcatgaacttgg 1800  
Db 10000 acttacactatttttagtttgcctctagtttatgaatgcatgtatgcatgaacttgg 10059  
Qy 1801 gagatatttttctcccaactcttcttcccaatttaaatgtgctgtctttagaagcca 1860  
Db 10060 gagatatttttctcccaactcttcttcccaatttaaatgtgctgtctttagaagcca 10119

Qy 1861 ctgctcagcttctgagctcagatcaccaaaaggaagctctgtgtacacagcatgataaaaga 1920  
Db 10120 ctgctcagcttctgagctcagatcaccaaaaggaagctctgtgtacacagcatgataaaaga 10179  
Qy 1921 caatgggacggggtcacagtgagctcccttcccttccaggggtatggagacgagctgtaga 1980  
Db 10180 caatgggacggggtcacagtgagctcccttcccttccaggggtatggagacgagctgtaga 10239  
Qy 1981 gagatgtctccagggaggttttcatttaacagcaatttagtcagatctgtgcatcctatgc 2040  
Db 10240 gagatgtctccagggaggttttcatttaacagcaatttagtcagatctgtgcatcctatgc 10299  
Qy 2041 tttaacagaaatgtcagtgaggcctgagatcatcagatgaggttcatcggttttcaatgt 2100  
Db 10300 tttaacagaaatgtcagtgaggcctgagatcatcagatgaggttcatcggttttcaatgt 10359  
Qy 2101 cccgtatccttttgaagaccttgaagttgcaacgcgggaaaaacaggaactccacctg 2160  
Db 10360 cccgtatccttttgaagaccttgaagttgcaacgcgggaaaaacaggaactccacctg 10419  
Qy 2161 gtgcgtaaatgacagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2220  
Db 10420 gtgcgtaaatgacagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 10479  
Qy 2221 tgacagagcttgtgaaccttttaactgggactgggcaagtgcaatccacacctttatacaat 2280  
Db 10480 tgacagagcttgtgaaccttttaactgggactgggcaagtgcaatccacacctttatacaat 10539  
Qy 2281 gaattgctgaagagggcctttt 2301  
Db 10540 gaattgctgaagagggcctttt 10560

RESULT 2  
C63529/c  
ID C63529 standard; DNA; 5513 BP.  
XX C63529;  
XX AC  
XX XX  
XX XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Slit protein coding sequence.  
XX DE  
XX DE  
XX Slit protein; frog; cell migration; neuronal; nerve axon; dendrite;  
XX KW leukocyte; graft rejection; HIV infection; antinflammatory;  
XX KW wound repair; organ regeneration; asthma; arthritis; glomerulonephritis;  
XX KW cystic fibrosis; ulcerative colitis; Crohn's disease; multiple sclerosis;  
XX KW allergic encephalomyelitis; Alzheimer's disease;  
XX KW coronary artery restenosis; ss.  
XX OS Xenopus sp.  
XX PN WO200055321-A2.  
XX XX  
XX PD 21-SEP-2000.  
XX PF  
XX PF 16-MAR-2000; 2000WO-US07040.  
XX XX  
XX PR 17-MAR-1999; 99US-0124767.  
XX XX  
XX PA (UNIW ) UNIV WASHINGTON.  
XX  
XX PI Rao Y, Wu JY;  
XX DR WPI; 2000-628199/60.  
XX DR P-PSDB; B28151.  
XX XX  
XX PT New vertebrate DNA encoding a slit protein that is useful for treating  
XX PT graft rejection, inhibiting infection of a cell by human  
XX PT immunodeficiency virus, inhibiting inflammation, and for aiding wound  
XX PT repair and organ regeneration -  
XX XX

PS Claim 1; Fig 1; 72pp; English.

XX The present sequence is the coding sequence for Xenopus slit protein. The  
 CC slit protein is useful for altering or guiding cell migration of neural  
 CC cells preferably nerve axons or dendrites, leukocytes or malignant cells  
 CC expressing roundabout (robo) protein, by repulsion or inhibition. Also,  
 CC the slit protein is useful for treating graft rejection, inhibiting the  
 CC infection of a cell by HIV, inhibiting inflammation and in aiding wound  
 CC repair and organ regeneration. In addition, the slit protein is useful  
 CC for the treatment of conditions involving the migration of leukocytes,  
 CC such as asthma, arthritis, glomerulonephritis, cystic fibrosis,  
 CC ulcerative colitis, Crohn's disease, multiple sclerosis, allergic  
 CC encephalomyelitis, Alzheimer's disease and coronary artery restenosis.

XX Sequence 5513 BP; 1668 A; 1216 C; 1211 G; 1418 T; 0 other;

Query Match 2.4%; Score 55.4; DB 21; Length 5513;  
 Best Local Similarity 57.1%; Pred. No. 3.4e-05;  
 Matches 101; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1669 ctccagcttgattggcatgagctcagagctttgattgaatgagttggagcccttagct 1728

Db 4114 CTGCAGAGTGATGTGGCTGAGCGGATCTTTGTCGAGGAATTGGGATAACGAGTC 4055

QY 1729 attgctcattagactacatttttagttttgctctgagtttatgaatatgcatgtat 1788

Db 4054 TTTGTTCCAAAAATTTAGCGTGAAGTATTTTCACATTTGTTTCACGATAGCCAGGCAG 3995

QY 1789 gcatgaactggagatattttctcccaatctcttctccatttaaatgtgc 1845

Db 3994 GCACTGGCAGAGATCATTCCTTTCACATGATGGGCTCCATTTTGACATTC 3938

RESULT 3

A07587/C  
 ID A07587 standard; DNA; 4064 BP.

XX A07587;

XX 29-AUG-2000 (first entry)

XX Mouse piwi gene, designated miwi.

XX Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;  
 KW anaemia; immunodeficiency; male infertility; mouse; ds.  
 XX Mus sp.

PH Key Location/Qualifiers  
 FT CDS 191..2779  
 FT /\*tag= a  
 FT /product= miwi  
 FT /transl\_except= (pos:450..452; aa:Xaa)  
 FT /transl\_except= (pos:1337..1339; aa:Xaa)  
 FT /transl\_except= (pos:2636..2638; aa:Xaa)  
 FT /note= "Xaa= leu or ile"  
 FT /transl\_except= (pos:836..838; aa:Xaa)  
 FT /note= "Xaa= unspecified amino acid"

XX WO200032039-A1.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-US28764.

XX 04-DEC-1998; 98US-0110901.

XX (UYDU-) UNIV DUKE.

XX Lin H;

XX WPI; 2000-412085/35.

XX

XX

XX

XX

XX

XX

XX

DR P-PSDB; Y90234.

XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene  
 CC therapy of diseases such as cancer and in various research and  
 CC diagnostic applications -  
 XX Claim 19; Page 180-185; 201pp; English.

XX This sequence encodes the mouse piwi family protein, designated  
 CC miwi. The piwi family nucleic acids and polypeptides are used in gene  
 CC therapy of diseases such as cancer and also in various research and  
 CC diagnostic applications. The sequences can also be used to treat  
 CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.

XX Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;

Query Match 2.0%; Score 46; DB 21; Length 4064;  
 Best Local Similarity 63.8%; Pred. No. 0.013;  
 Matches 70; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1374 ttagtattctctctctgagacagagctcactgtgtgcccaggtctcctcaact 1433

Db 3656 TTCTTTTCTCTGTTGTTAAGACAAAGTCTCACTATGTAGCCTAGCTGACCGGAACT 3597

QY 1434 tgcggtccattttgtcctcactcagaaatgctgggcttccagggtgtgtgc 1483

Db 3596 TGGGGTTCCTGGCTTCACCTCCTGAGTCTGAGATCACAGGCGTGTGC 3547

RESULT 4

X83007  
 ID X83007 standard; DNA; 51259 BP.

XX X83007;

XX 31-AUG-1999 (first entry)

XX Partial mouse WRN genomic sequence #3.

XX Mouse; WRN; Werner's syndrome; detection; diagnosis; autosomal;  
 KW recessive disorder; phenotype; ss.  
 XX Mus musculus.

XX WO9724435-A1.

XX 10-JUL-1997.

XX 30-DEC-1996; 96WO-US20785.

XX 12-APR-1996; 96US-0632175.

XX 29-DEC-1995; 95US-0009409.

XX 29-DEC-1995; 95US-0580539.

XX 30-JAN-1996; 96US-0010835.

XX 30-JAN-1996; 96US-0594242.

XX (DARW-) DARWIN MOLECULAR CORP.

XX (OSHI/) OSHIMA J.

XX Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;

XX WPI; 1997-363671/33.

XX Isolated nucleic acid molecule encoding the WRN gene product -  
 CC useful for detection and treatment of Werner's syndrome, and related  
 CC diseases  
 XX Claim 1; Fig 7; 153pp; English.

XX This sequence represents a fragment of the genomic sequence containing  
 CC the coding region for the mouse WRN gene (X83004). The corresponding  
 CC human gene (X83001) encodes a protein related to Werner's syndrome.









PD 25-FEB-1999.  
 XX 18-AUG-1998; 98WO-US17044.  
 XX 16-JUN-1998; 98US-0092956.  
 PR 15-JUL-1998; 98US-0092956.  
 PR 19-AUG-1997; 97US-0056368.  
 PR 19-AUG-1997; 97US-0056369.  
 PR 19-AUG-1997; 97US-0056535.  
 PR 19-AUG-1997; 97US-0056555.  
 PR 19-AUG-1997; 97US-0056556.  
 PR 19-AUG-1997; 97US-0056628.  
 PR 19-AUG-1997; 97US-0056629.  
 PR 19-AUG-1997; 97US-0056726.  
 PR 19-AUG-1997; 97US-0056728.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Brewer LA, Duan R, Ebner R, Endress GA, Feng P;  
 PI Florence C, Florence KA, Komatsoulis GA, Laflaur DW;  
 PI Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR;  
 PI Young PE;  
 XX WPI: 1999-190160/16.  
 DR P-PSDB; Y07808.  
 XX New isolated human genes and the secreted polypeptides they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation of blood disorders  
 XX Claim 1a; Page 226-227; 280pp; English.  
 XX This invention describes novel isolated human secreted proteins and  
 CC their encoding nucleic acid sequences. The products of the invention  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also pathological conditions can be  
 CC diagnosed by determining the presence or amount of expression of  
 CC the new polypeptides in a sample or by determining the presence or  
 CC absence of mutations in the new polynucleotides. Specific uses are  
 CC described for each of the 70 polynucleotides, based on which tissues they  
 CC are most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, tumours, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, leukemias, diseases of the immune system, autoimmune diseases,  
 CC hepatic and renal disease, lymphomas, inflammation, allergies, asthma,  
 CC sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,  
 CC osteoporosis, arthritis, psoriasis, digestive/endocrine disorders,  
 CC infections and AIDS. The human secreted proteins of the invention are  
 CC represented in Y07744-Y07850 and the encoding nucleic acids are  
 CC represented in X37369-X37441.  
 XX Sequence 687 BP; 187 A; 148 C; 161 G; 187 T; 4 other;  
 SQ

Query Match 1.8%; Score 42; DB 20; Length 687;  
 Best Local Similarity 69.5%; Pred. No. 0.068;  
 Matches 57; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 1359 cctgcctgattaatagttattcttctctctgagacagagtcactgtggccca 1418  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 670 CCCCCCTCGAGTGTGTTTTTTTTTTTTTTTTTTTATAGACAGGGGCTCTACTATGTTGCCCA 611  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1419 ggctagtctcaaacctgcggtc 1440  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 610 GGCTGGTCTCAACTCTAGGCG 589  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13  
 C98178/c  
 ID C98178 standard; cDNA; 833 BP.  
 XX C98178;  
 AC C98178;  
 XX

DT 09-MAR-2001 (first entry)  
 XX Human colon cancer antigen nucleotide sequence SEQ ID NO:188.  
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO:188.  
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder; ss.  
 XX Homo sapiens.  
 OS WO200055351-A1.  
 XX PN 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US05883.  
 XX 12-MAR-1999; 99US-0124270.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Ruben SM;  
 XX WPI: 2000-587534/55.  
 XX P-PSDB; B53421.  
 DR Colon cancer associated gene sequences, referred to as colon cancer  
 DR antigens, useful for the treatment, prevention, and diagnosis of colon  
 DR disorders such as colon cancer -  
 PT Claim 1; Page 611-612; 2104pp; English.  
 XX C97991 to C98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in B53234 to B54006. The human  
 CC colon cancer antigens can have cytostatic, cardioactive, muscular;  
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and  
 CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins  
 CC may also be used to prevent diseases such as neural disorders, immune  
 CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, wounds, renal disorders, infectious  
 CC diseases, and cardiovascular disorders. C98764 to C98772 and B54007  
 CC represent sequences used in the exemplification of the present  
 CC invention.  
 XX Sequence 833 BP; 205 A; 212 C; 164 G; 247 T; 5 other;  
 SQ

Query Match 1.8%; Score 41.8; DB 21; Length 833;  
 Best Local Similarity 67.4%; Pred. No. 0.086;  
 Matches 58; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1343 ttagaagcctggatgcctgcctattatatttcttctctgagacagagt 1402  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 127 TTATAAGCCAGGGTACCTCAGGAGGTTCTATTACTATTATTATTATTGAGACAGACT 68  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1403 ctcaactgtgtgcccaggctagctctc 1428  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 67 CTNACTGTGTCGCCCCAGGCAGATCTC 42

RESULT 14  
 V59610/c  
 ID V59610 standard; DNA; 1751 BP.  
 XX V59610;  
 AC V59610;

XX 06-JAN-1999 (first entry)  
 XX Human secreted protein gene 100 clone HLQAB52.  
 DE  
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergic; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO9839448-A2.  
 XX  
 XX 11-SEP-1998.  
 XX  
 XX 06-MAR-1998; 98WO-US04493.  
 XX  
 XX 02-OCT-1997; 97US-0061060.  
 PR 23-MAY-1997; 97US-0047615.  
 PR 23-MAY-1997; 97US-0047617.  
 PR 23-MAY-1997; 97US-0047618.  
 PR 23-MAY-1997; 97US-0047632.  
 PR 23-MAY-1997; 97US-0047633.  
 PR 06-JUN-1997; 97US-0048964.  
 PR 06-JUN-1997; 97US-0048974.  
 PR 13-JUN-1997; 97US-0049610.  
 PR 08-JUL-1997; 97US-0051926.  
 PR 16-JUL-1997; 97US-0052874.  
 PR 18-AUG-1997; 97US-0055724.  
 PR 22-AUG-1997; 97US-0056630.  
 PR 22-AUG-1997; 97US-0056631.  
 PR 22-AUG-1997; 97US-0056632.  
 PR 22-AUG-1997; 97US-0056636.  
 PR 22-AUG-1997; 97US-0056637.  
 PR 22-AUG-1997; 97US-0056662.  
 PR 22-AUG-1997; 97US-0056664.  
 PR 22-AUG-1997; 97US-0056845.  
 PR 22-AUG-1997; 97US-0056862.  
 PR 22-AUG-1997; 97US-0056864.  
 PR 22-AUG-1997; 97US-0056872.  
 PR 22-AUG-1997; 97US-0056874.  
 PR 22-AUG-1997; 97US-0056875.  
 PR 22-AUG-1997; 97US-0056876.  
 PR 22-AUG-1997; 97US-0056877.  
 PR 22-AUG-1997; 97US-0056878.  
 PR 22-AUG-1997; 97US-0056879.  
 PR 22-AUG-1997; 97US-0056880.  
 PR 22-AUG-1997; 97US-0056881.  
 PR 22-AUG-1997; 97US-0056882.  
 PR 22-AUG-1997; 97US-0056884.  
 PR 22-AUG-1997; 97US-0056886.  
 PR 22-AUG-1997; 97US-0056887.  
 PR 22-AUG-1997; 97US-0056888.  
 PR 22-AUG-1997; 97US-0056889.  
 PR 22-AUG-1997; 97US-0056892.  
 PR 22-AUG-1997; 97US-0056893.  
 PR 22-AUG-1997; 97US-0056894.  
 PR 22-AUG-1997; 97US-0056903.  
 PR 22-AUG-1997; 97US-0056908.  
 PR 22-AUG-1997; 97US-0056909.  
 PR 22-AUG-1997; 97US-0056910.  
 PR 22-AUG-1997; 97US-0056911.  
 PR 05-SEP-1997; 97US-0057650.  
 PR 05-SEP-1997; 97US-0057669.  
 PR 05-SEP-1997; 97US-0057761.  
 PR 12-SEP-1997; 97US-0058785.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Bedharik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;  
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX  
 XX WPI; 1998-506364/43.  
 DR P-PSDB; W74830.  
 XX  
 XX New isolated human genes and the secreted polypeptide(s) they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Claim 1; Page 333-334; 721pp; English.  
 XX  
 XX This sequence represents a nucleic acid molecule designated Gene 101 from  
 CC the human cDNA clone HLQAB52 (deposited as clone ATCC 97901 and ATCC  
 CC 209047) which encodes a secreted human protein. The gene can be used to  
 CC generate fusion proteins by linking to the gene to a human immunoglobulin  
 CC Fc portion (e.g. V59502) for increasing the stability of the fused  
 CC protein as compared to the human protein only.  
 CC The invention relates to 186 novel genes and their fragments (nucleic  
 CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 13:30:07 ; Search time 150.46 Seconds  
(without alignments)  
2833.205 Million cell updates/sec

Title: US-09-445-201-1-copy\_8260\_10560

Perfect score: 2301

Sequence: 1 tgaataagatgaggtgcc.....aattgctgaagagcctttt 2301

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.6	2.1	7218	1	US-08-232-463-14
2	45.6	2.0	51259	3	US-08-781-891-209
3	43.2	1.9	152331	4	US-09-128-155-16
4	42.2	1.8	14507	3	US-08-785-150-1
5	40.8	1.8	152331	4	US-09-128-155-16
6	40.4	1.8	176373	4	US-09-128-155-17
7	40.2	1.7	380	1	US-08-126-587C-5
8	40	1.7	1920	1	US-08-087-772A-1
9	39.6	1.7	246240	2	US-08-724-394A-20
10	39.6	1.7	246240	2	US-08-724-394A-21
11	39.6	1.7	246240	2	US-08-724-394A-22
12	39.4	1.7	2880	4	US-09-115-954-3
13	39.4	1.7	3842	4	US-09-115-954-7
14	39.4	1.7	3912	4	US-09-115-954-1
15	39.2	1.7	11725	2	US-08-756-506-1
16	39.2	1.7	17041	1	US-08-076-011-1
17	39	1.7	35060	3	US-08-814-095-7
18	38.6	1.7	2908	3	US-08-487-799-1
19	38.6	1.7	6769	1	US-08-480-784-20
20	38.6	1.7	6769	1	US-08-483-553-20
21	38.6	1.7	6769	1	US-08-487-002-20
22	38.6	1.7	6769	1	US-08-483-554B-20
23	38.6	1.7	6769	1	US-08-488-011B-20
24	38.6	1.7	6769	4	US-08-850-727-20
25	38.6	1.7	6769	5	PCT-US95-10202-20
26	38.6	1.7	6769	5	PCT-US95-10203-20
27	38.6	1.7	6769	5	PCT-US95-10220-20

c 28 38.4 1.7 501 3 US-08-699-628-1  
 29 38.4 1.7 2520 2 US-08-454-557C-50  
 30 38.4 1.7 2520 2 US-08-340-426D-50  
 31 38.4 1.7 2520 2 US-08-450-673C-50  
 32 38.4 1.7 2520 5 PCT-US95-17111A-50  
 c 33 38.2 1.7 5095 1 US-08-092-817-3  
 34 38 1.7 31571 1 US-08-323-443B-1  
 35 38 1.7 40352 3 US-08-846-111D-15  
 36 38 1.7 53526 3 US-08-658-136-2  
 37 38 1.7 53577 3 US-08-658-136-1  
 38 37.6 1.6 615 4 US-08-998-416-186  
 39 37.4 1.6 7720 4 US-09-318-448-5  
 40 37 1.6 4084 3 US-08-866-340-1  
 41 37 1.6 4460 4 US-09-103-875-4  
 42 36.8 1.6 4922 2 US-08-330-272-5  
 43 36.8 1.6 4922 5 PCT-US95-13663-5  
 c 44 36.6 1.6 5375 4 US-08-757-223-7  
 c 45 36.6 1.6 10825 3 US-08-652-265-1

#### ALIGNMENTS

RESULT 1  
 US-08-232-463-14  
 ; Sequence 14, Application US/08232463  
 ; Patent No. 5670367  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DORNER, F.  
 ; APPLICANT: SCHEIFLINGER, F.  
 ; APPLICANT: FALKNER, F. G.  
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 ; NUMBER OF SEQUENCES: 52  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 1800 Diagonal Road, Suite 500  
 ; CITY: Alexandria  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22313-0299  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/232,463  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/935,313  
 ; FILING DATE:  
 ; APPLICATION NUMBER: EP 91 114 300.6  
 ; FILING DATE: 26-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)836-9300  
 ; TELEFAX: (703)683-4109  
 ; TELEX: 899149  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 7218 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; CLONE: pT29pt-Fls  
 ; US-08-232-463-14

```

; INFORMATION FOR SEQ ID NO: 209:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 51259 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
US-08-781-891-209

Query Match                2.0%;   Score 45.6;   DB 3;   Length 51259;
Best Local Similarity      62.1%;   Pred. No. 0.009;
Matches 72; Conservative  0; Mismatches 44; Indels 0; Gaps

QY  1368  tattaattagttattctctctctgagacagagtgctcactgtgtgcccaggttagtct 1427
      |||||  ||  |||  |||||  |||||  |||||  ||  ||  ||
Db  23743  TTTTAGTTTGTAGTTTACTTTTTTTTAGACAGGGTCTCACTGTGTAGCTGGGACAAGCT 23802
      |||||  ||  |||  |||||  |||||  |||||  ||  ||  ||

QY  1428  caaactgcggtccattgtctcactcactcactcagaatgctgggcttcagtggtgtgc 1483
      |||||  ||  ||||  |||||  ||  |||||  ||  |||||  ||  |||||
Db  23803  CCACCCCTGTGCCCTTTTCTCACCCTCTGAGTCTGGGATCACAGGCGTGTGC 23858
      |||||  ||  ||||  |||||  ||  |||||  ||  |||||  ||  |||||

RESULT 3
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

[illegible]

RESULT	5	
US-09-128-155-16/c		
; Sequence 16, Application US/09128155		
; Patent No. 6117654		
; GENERAL INFORMATION:		
; APPLICANT: Pan, Yang		
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY		
; TITLE OF INVENTION: AND USES THEREOF		
; FILE REFERENCE: 09404/052001		
; CURRENT APPLICATION NUMBER: US/09/128,155		
; CURRENT FILING DATE: 1998-08-03		
; EARLIER APPLICATION NUMBER: US 60/091,650		
; EARLIER FILING DATE: 1998-07-02		

```

RESULT      6
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

```

[illegible]

RESULT 7  
US-08-126-587C-5/c  
; Sequence 5, Application US/08126587C  
; Patent No 5534438

```

; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Goldberg, Paul
; APPLICANT: Andrew, Susan
; APPLICANT: Rommens, Johanna M.
; APPLICANT: Lin, Blaovang
; TITLE OF INVENTION: Process for Isolating Genes and the Gene
; TITLE OF INVENTION: Causative of Huntington's Disease and Differential 3'
; TITLE OF INVENTION: Polyadenylation in the Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer Park & Gibson
; STREET: 1211 E. Morehead Street
; CITY: Charlotte
; STATE: NO. 5534438th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/126,587C
; FILING DATE: 24-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Layton Jr., Samuel G.
; REGISTRATION NUMBER: 22,807
; REFERENCE/DOCKET NUMBER: 3477-84
; TELEPHONE: 704-377-1561
; TELEFAX: 704-334-2014
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4p16.3
; UNITS: bp
; US-08-126-587C-5

Query Match 1.7%; Score 40.2; DB 1; Length 380;
Best Local Similarity 60.6%; Pred. No. 0.019;
Matches 66; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1368 tattaattagttattcttctgtgacagagtgctcactgtgtgcccaggctagtct 1427
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1428 caaactggcggtccattgtctcactcactcagaatgtgggtccagg 1476
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 279 GCGGACTGCAGTGGCGCAATCTCGGCTCACTGAAAGCTCCGCTTCCCGG 231
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-08-087-772A-1/C
; Sequence 1, Application US/08087772A
; Patent No. 5691155
; GENERAL INFORMATION:
; APPLICANT: Namias, Clara
; APPLICANT: Emorine, Jean L.
; APPLICANT: Strosberg, Donny A.
; TITLE OF INVENTION: Nucleotide Sequences Encoding the Murine
; TITLE OF INVENTION: Beta3-Adrenergic Receptor and Their Applications
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Bell, Seltzer, Park & Gibson
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5691155th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,772A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Linker, Raymond O.
; REGISTRATION NUMBER: 26,419
; REFERENCE/DOCKET NUMBER: 3339-195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 568..1731
; US-08-087-772A-1

Query Match 1.7%; Score 40; DB 1; Length 1920;
Best Local Similarity 63.5%; Pred. No. 0.058;
Matches 61; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1382 tcttctcttgagacagagtgctcactgtgtgcccaggctagtctcaaaacttgcgtcc 1441
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 97 TCTATTGTTGAGATAGTCTTCTATATGTTGCCAGATTTGCTGGACCTCCAGATCC 38
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1442 attgtctcactcactcagaatgtgggtccagg 1477
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 37 TTCTGCTCAACCTCCCGAGTCTGGATTACAGAT 2
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



GENERAL INFORMATION:  
 APPLICANT: Feder, John N.  
 APPLICANT: Kronmal, Gregory S.  
 APPLICANT: Lauer, Peter M.  
 APPLICANT: Ruddy, David A.  
 APPLICANT: Thomas, Winston  
 APPLICANT: Tsuchihashi, Zenta  
 APPLICANT: Wolff, Roger K.  
 TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
 TITLE OF INVENTION: Sequences and Antibodies Thereeto  
 NUMBER OF SEQUENCES: 31  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
 STREET: Two Embarcadero Center, 8th Floor  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/724,394A  
 FILING DATE: 01-OCT-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fitts, Renee A.  
 REGISTRATION NUMBER: 35,136  
 REFERENCE/DOCKET NUMBER: 017957-000100  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-576-0200  
 TELEFAX: 415-576-0300  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 246240 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match
Best Local Similarity 1.7%; Score 39.6; DB 2; Length 246240;
Matches 54; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1363 cctgtattattagttattctctctgtgagacagagtcactgtgtgcccaggtc 1422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95222 CCITTTTTTTTTTCTCTCTCTTTTGTAGACAGGGCTCACTATGTTGCCAGGCT 95281

QY 1423 agtctcaaaactgcggtc 1440
||||| ||||| |||
Db 95282 GGTCTCAACTTCTGGGC 95299

RESULT 12
US-09-115-954-3/c
; Sequence 3, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2880
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-3

Query Match
Best Local Similarity 1.7%; Score 39.4; DB 4; Length 2880;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1368 tattaattagttattctctctgtgagacagagtcactgtgtgcccaggtc 1427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2853 TTTTFTTTTTTTTTTTTTTTTGTAGACAGAGGGTCTCATTTATGCTAGGCTAGTCT 2794

QY 1428 caaactgcggtc 1440
||||| ||||| |||
Db 2793 CAAACTTCTGGCC 2781

RESULT 13
US-09-115-954-7/c
; Sequence 7, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
```

```
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 3842
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-7

Query Match
Best Local Similarity 1.7%; Score 39.4; DB 4; Length 3842;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1368 tattaattagttattctctctgtgagacagagtcactgtgtgcccaggtc 1427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3814 TTTTFTTTTTTTTTTTTTTTTGTAGACAGAGGGTCTCATTTATGCTAGGCTAGTCT 3755

QY 1428 caaactgcggtc 1440
||||| ||||| |||
Db 3754 CAAACTTCTGGCC 3742

RESULT 14
US-09-115-954-1/c
; Sequence 1, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3912
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-115-954-1

Query Match
Best Local Similarity 1.7%; Score 39.4; DB 4; Length 3912;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1368 tattaattagttattctctctgtgagacagagtcactgtgtgcccaggtc 1427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3884 TTTTFTTTTTTTTTTTTTTTTGTAGACAGAGGGTCTCATTTATGCTAGGCTAGTCT 3825

QY 1428 caaactgcggtc 1440
||||| ||||| |||
Db 3824 CAAACTTCTGGCC 3812

RESULT 15
US-08-756-506-1/c
```

```
; Sequence 1, Application US/08756506
; Patent No. 5905185
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; TITLE OF INVENTION: ANIMALS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,506
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 95-28
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11725 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(3520..3630, 5093..5117, 5210..5347, 5450
; LOCATION: ..5584, 8253..8395, 9269..9386, 10516..11102)
; US-08-756-506-1
```

```
Query Match 1.7%; Score 39.2; DB 2; Length 11725;
Best Local Similarity 73.5%; Pred. No. 0.3;
Matches 50; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1366 gctattaattagttattctctctctctgagacagagtcctcactgtgtgccccaggctagt 1425
   ||||| || || || || || || || || || || || || || || || || || || || ||
Db 6868 GATCTTTTTCATCTTTTAAAAAATAGACACAGGGTCTCACTGTGTGCCAGGCTAGT 6809

QY 1426 ctcaaaact 1433
   |||||
Db 6808 CTCAAACT 6801
```

Search completed: June 21, 2001, 18:05:44  
Job time: 16537 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 13:26:36 ; Search time 4579.66 Seconds  
(without alignments)  
4749.483 Million cell updates/sec

Title: US-09-445-201-1-copy\_8260\_10560  
Perfect score: 2301  
Sequence: 1 tgaataagatgaggtgcc.....aattgctgaagagcctttt 2301

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues  
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

em\_esthum10:\*  
44: em\_esthum11:\*  
45: em\_esthum12:\*  
46: em\_esthum13:\*  
47: em\_esthum14:\*  
48: em\_esthum15:\*  
49: em\_esthum16:\*  
50: em\_esthum17:\*  
51: em\_esthum18:\*  
52: em\_esthum19:\*  
53: em\_esthum20:\*  
54: em\_esthum21:\*  
55: em\_esthum22:\*  
56: em\_esthum23:\*  
57: em\_esthum24:\*  
58: em\_esthum25:\*  
59: em\_esthum26:\*  
60: em\_esthum27:\*  
61: em\_esthum28:\*  
62: em\_estin1:\*  
63: em\_estin2:\*  
64: em\_estin3:\*  
65: em\_estin4:\*  
66: em\_estin5:\*  
67: em\_estom1:\*  
68: em\_estom2:\*  
69: em\_estov1:\*  
70: em\_estov2:\*  
71: em\_estpl1:\*  
72: em\_estpl2:\*  
73: em\_estpl3:\*  
74: em\_estpl4:\*  
75: em\_estpl5:\*  
76: em\_estpl6:\*  
77: em\_estpl7:\*  
78: em\_estpl8:\*  
79: em\_estpl9:\*  
80: em\_estpl10:\*  
81: em\_estro1:\*  
82: em\_estro2:\*  
83: em\_estro3:\*  
84: em\_estro4:\*  
85: em\_estro5:\*  
86: em\_estro6:\*  
87: em\_estro7:\*  
88: em\_estro8:\*  
89: em\_estro9:\*  
90: em\_estro10:\*  
91: em\_estro11:\*  
92: em\_estro12:\*  
93: em\_estro13:\*  
94: em\_estro14:\*  
95: em\_estro15:\*  
96: em\_estro16:\*  
97: em\_estro17:\*  
98: em\_estro18:\*  
99: em\_estro19:\*  
100: em\_estro20:\*  
101: gb\_est25:\*  
102: gb\_est26:\*  
103: gb\_est27:\*  
104: gb\_est28:\*  
105: gb\_est29:\*  
106: gb\_est30:\*  
107: gb\_est31:\*  
108: gb\_est32:\*  
109: gb\_est33:\*  
110: gb\_est34:\*  
111: gb\_est35:\*  
112: gb\_est36:\*  
113: gb\_est37:\*  
114: gb\_est38:\*  
115: gb\_est39:\*  
116: gb\_est40:\*  
117: gb\_est41:\*  
118: gb\_est42:\*  
119: gb\_est43:\*  
120: gb\_est44:\*  
121: gb\_est45:\*  
122: gb\_est46:\*  
123: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



```

Db 300 CTTCTCTCTGACAGAGTCTCACTGTGTGGCCAGGCTAGTCTCAAACTTGCCTGCTCA 359
QY 1443 ttgtctcaactcatcagaatgctggctccagggtgtgtgaccacactagtagctgc 1502
Db 360 TTTGCTCTCACTCATCAGAACTGCTGGCTTCCAGGTTGTGACACACACTAGTAGCTGCG 419
QY 1503 gtttaagctaaagctggaagatcctcatgctcccttaccatggtgggaatgttacagt 1562
Db 420 GTTTTAAGCTAAGAGCTGAAGATCCTCATGCTCTCTTTTACCAGTGTGGGCTATTACAGT 479
QY 1563 tagttgactgaactagttatcctgctgtatgaatgaactgaggtggtatgtatctctca 1622
Db 480 TAGTTGACTGAAGAACTAGTTATCTGCTGTGTGAATGACCTGAGTGGTATGATCTCTCA 539
QY 1623 agatgcttttttgcatttcaatcagttaggttaacaaagt-tcttaagctctcagactgtgta 1681
Db 540 AGATGCTTTTTTGGCATTTCAATCAGTAGTAGTAAACAAGTGTCTTATGTCTCCAGCTTTGTA 599
QY 1682 ttggcagtcagctgagctgttgatgaatagttgagtggtggacccctagctattgctcattaga 1741
Db 600 TTGCTATGAGCTCAGAGCTTTGATTAATGAGTTGGGACCCCTAGCTATTGCTCAATTAGA 659
QY 1742 ctacacatttttagtttgcctgagtttataaataatgcatgcatgataaacttggg 1801
Db -660 CTTACACTATTTTGTCTCTGAGTTTATGAATATGATGATGATGATGATGATGATGATGATG 719
QY 1802 agatatttttctcccaattccctttt 1828
Db 720 AGATATTCTCTTCCCAATTCCTTTT 746

```

RESULT 2

```

AZ364007 586 bp DNA GSS 02-OCT-2000
LOCUS LM0110F03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0110F03 F, DNA sequence.
ACCESSION AZ364007
VERSION AZ364007.1 GI:10477707
SOURCE GSS.
ORGANISM house mouse.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 586)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0110 row: F column: 03
Seq primer: CGTTGTAACACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 586.
Location/Qualifiers
1. .586
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0110F03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"

```

FEATURES

Source

```

/lab_host="E: Coli strain XL10-Gold, Tl-resistant, F-
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 137 a 125 c 137 g 186 t 1 others
ORIGIN

```

Query Match 24.4%; Score 560.8; DB 242; Length 586;  
 Best Local Similarity 98.5%; Pred. No. 2.7e-135;  
 Matches 576; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

```

QY 1414 gccaggcagtcctcaaaactgcggtccatttgcctcaactcagatgctgggctcc 1473
Db 2 GGGCCAGCTAGTCTCAAACTTGGCGTCCATTGTCTCACTCATCAGAAATGCTGGGCTTCC 61
QY 1474 aggtgtgtgcaccacactaggtagctgcgtctttaagctaaagctggaagatcctgatg 1533
Db 62 AGGTGTGTGCACACACTAGGTAGCTCGCGTTTAACTAAGAGCTGGAGATCCTGATG 121
QY 1534 tcccttaccatggtgggcatgttacaggttagttgactgaaactagttatctcgtctgtg 1593
Db 122 TCGTTTACCATGTTGGGGCATGTTACAGGTAGTTGACTGAAACTAGTTATCTCGCTGTG 181
QY 1594 taatgacctgagtgatgtatctctcaagatgctttttgcatttcaatcagttaggt 1653
Db 182 TAATGACCTGCTAGTGTGTATCTCTCAAGATGCTTTTTCATTTCAATCAGTTAGGT 241
QY 1654 acaaat-tcttaagtctccagcttggtattggcatgagctcagagcttttgatattagag 1712
Db 242 AACAAAGTGTCTTATGCTCCAGCTTTGTATGTTGATGAGCTCAGAGCTTGTATTAATGAG 301
QY 1713 ttgggacccctagctattgctcattagactacactatttttagttttgctctgagttt 1772
Db 302 TTGGGACCCCTAGCTATTGCTCATTAGACTTACACTATTATTTTAGTTTGTCTGAGTTT 361
QY 1773 atgaatatcatgtatgcatgaacttggagatatatttctcccaattccttttctc 1832
Db 362 ATGAATATCATGTATGCTGATGAACCTTGGGAGATATTTCCTCTCCCAATTTCTTTTCTC 421
QY 1833 catttaaatgtctgtcttttagaagcactgctcagcttctcagctcagataccacaaag 1892
Db 422 CATTTAAATGTCTGCTTTTAGAAGCCACTGCTCAGCTTCTGAGCTCAGATACCAAG 481
QY 1893 gaagtctgtacacagcatgataaaagacaatggcggtggtcacagtggtccctcc 1952
Db 482 GAAGTCTGTGTACACAGCATGATAAAAGACAAATGGGACGNGTACAGTGTGCTCCGTC 541
QY 1953 ttccaggggtatggagacagctgttagagatgtctccagggag 1997
Db 542 TTTCAGGGGTATGGACGAGCTGTAGAGAGATGTCTCCAGGGAG 586

```

RESULT 3

LOCUS AQ126020/c

DEFINITION HS\_3032\_B2\_C02\_MF CIT Approved Human Genomic Sperm Library D Homo

LOCUS AQ126020 404 bp DNA GSS 23-SEP-1998



Accession		Version		Keywords		Source		Organism		Reference		Authors		Title		Journal		Medline		Comment												
sapiens genomic clone Plate=3032 Col=4 Row=F, DNA sequence.		A0126020		A0126020.1 GI:3503186		GSS.		human.		Homo sapiens		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		Mahairs,G.G., Wallace,J.C., Smith,K., Swartzelli,S., Holzman,T., Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		99380589		Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3032 row: F column: 4 Class: BAC ends High quality sequence stop: 404.										
FEATURES		source		1. 404		/organism="Homo sapiens"		/db_xref="taxon:9606"		/clone="Plate=3032 Col=4 Row=F"		/clone_lib="CIT Approved Human Genomic Sperm Library D"		/sex="male"		/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"		160 a		69 c		83 g		92 t								
BASE COUNT		ORIGIN		Query Match		3.4%;		Score 79.2;		DB 224;		Length 404;		Best Local Similarity		56.8%;		Pred. No. 8.4e-10;		Matches 135;		Conservative 0;		Mismatches 53;		Indels 14;		Gaps 1;				
QY	1195	ccacagtg	ccacggtc	agcgttg	gctgctatt	gctgctt	ggctctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt	gctgctt		
Db	316	CCTCAGATTC	CAATGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT	CTTGGC	ACTGGAT		
QY	1247	-----tg	acatg	gcctcat	ctcttg	cttgc	gttcac	ttcact	attt	gccc	aaac	acgtc	aatgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	atgc	
Db	256	TCGGTCT	CATAC	TCTCG	CTCC	PATTTG	TTGTC	ATAC	TCTCT	CTCT	TATTTG	CCCA	CAAGG	TC	CAATGC	197																
QY	1301	cagctg	agc	gcttag	gagtc	atc	gttctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	ctt	
Db	196	CAGTCTAG	GCATAG	GGGAG	CTCAT	TTTTC	TTTCT	TAGT	TATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG
QY	1361	tgctg	ctct	ataa	ttag	ttatt	1382																									
Db	136	TGCCCA	CATC	CAAT	TACT	TTT	115																									
RESULT	4	A0332669/c		LOCUS		A0332669		464 bp		DNA		GSS		06-MAR-1999		HS_5004		AL_A09		SPFE		R8C11		Human Male		BAC Library		Homo sapiens		genomic clone Plate=380 Col=17 Row=A, DNA sequence.		
ACCESSION	A0332669	A0332669		VERSION		A0332669		GI:4341028		GSS.		human.		Homo sapiens		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		Mahairs,G.G., Wallace,J.C., Smith,K., Swartzelli,S., Holzman,T., Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		99380589		Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3032 row: F column: 4 Class: BAC ends High quality sequence stop: 404.						

1 (bases 1 to 464)  
Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
On Feb 19, 1999 this sequence version replaced gi:4130296.  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 580 row: A column: 17  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 464.  
Location/Qualifiers  
    1..464  
        /organism="Homo sapiens"  
        /db\_xref="taxon:9606"  
        /clone="plate=580 Col=17 Row=A"  
        /clone\_lib="RPC111 Human Male BAC Library"  
        /sex="Male"  
        /cell\_type="Lymphocytes"  
        /note="Vector: pBACE3.6; RPC111 Human Male BAC Library"  
BASE COUNT      178 a      75 c      113 t          5 others  
ORIGIN

Query Match                  3.4%; Score 78.6; DB:227; Length 464;  
Best Local Similarity      66.5%; Pred. No. 1.2e+09;  
Matches 125; Conservative   0; Mismatches 62; Indels 1; Gaps 1;

QY 1196 cacagtgcacgctaggctattgacctggcctgttgcttgaataactttat-tgacatgg 1254  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 298 CACGCGAATGTCAGCTGGGTATTTACTTATACCTCTCTATTATTTCCGCTCATACTT 239  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1255 cctcatcttgttcacgtctcaacttatgtgccacaacagctaatgccagctgagcgctta 1314  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 238 CGTCCATTTGTTTCATACTCTCTTATTATGGCCAGCAAGTCAATGCCAGTTTAGCCCTAG 179  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1315 ggagctactgttcttttagtcagtcgcaattagaagccctggaatgcctgcctgtattaa 1374  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 178 GGAGTCATATTTTCTTGTTAGTTGATATGACTTAGAAAGCTTGGGTGCNTGCCCCACATCAAT 119  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1375 tagttatt 1382  
|||||  
Db 118 TACTTTWT 111

RESULT 5  
AZ508851/c  
LOCUS  
DEFINITION IM0351IG04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0351G04 R, DNA sequence.  
ACCESSION AZ508851  
VERSION AZ508851.1 GI:10690167  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 556)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
.N., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choiri.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 67 row: J column: 15

Email: szhac@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 67 row: J column: 15  
 Fax: 301 838 0208

Query Match	2.5%	Score 56.6;	DB 248;	Length 506;
Best Local Similarity	60.9%;	Pred. No. 0.00069;		
Matches 109;	Conservative	0;	Mismatches 69;	Indels 11;
				Gaps 1;

1334	cagtcgcaattagaaagcctggatgcctgcctgctattaatagttattctctctctc	1393
QY		
324	CACCTGAGTCCAGCAACACATTGGTTGAAAATATCTCTTTAAAAAATATTTTTTTT	383
DB		

	1394	agacagagtctcactgtgtggccaggcgtagtctcaaaactgcggtccattgtctcat	1453
QY			
	384	AAAGAAGATCTTACTATGTAGCCACAGGTTGCCTTAAATGTGTGATCATCTACCCTAGT	443
Db			

[illegible]

RESULT 7  
RE651040

LOCUS	EST	06-SEP-2000
BE651040	256 bp	musculus cdna clone
UI-M-BH3-asy-e-09-0-UI.r1	NIH-BMAP_M_S4	
UI-M-BH3-asy-e-09-0-UI.5	5', mRNA sequence.	

ACCESSION	BE651040	
VERSION	BE651040.1	GI:9976864
KEYWORDS	EST,	
SOURCE	house mouse.	

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.



FEATURES	source
Location/Qualifiers	1. 559
	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="taxon:10090"
	/clone="UUGCLM020C02"
	/clone_lib="Mouse 10kb plasmid UUGCLM library"
	/sex="Male"
	/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"





available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 1-41, >AT\_richLow complexity 100-201, >PB1D10#SINE/Alu  
Seq primer: M13 Forward  
POLYA=yes.

## FEATURES

source Location/Qualifiers  
1. .431  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="NIH\_BMAP\_MSC"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH-BMAP\_MSC library is a non-normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.  
TAG\_LIB=NIH\_BMAP\_MSC  
TAG\_TISSUE=spinal-cord  
TAG\_SEQ=TCGAA"

BASE COUNT 97 a 105 c 84 g 145 t

## ORIGIN

Query Match 2.3%; Score 53.4; DB 102; Length 431;  
Best Local Similarity 69.9%; Pred. No. 0.0045;  
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
Qy 1385 tctcttgagacagagtcctcactgtgtgcccaggtagtctcaaaacttgcggtccatt 1444  
Db 102 TCTCTTTTGAGACAGGAACCTTGCTATGTAGCCAGGCTGGCTTTGAACTTGGGATCCTCT 161  
Qy 1445 tgtctcactcatcagaatgctgggtccagggtgtgcacca 1487  
Db 162 TGGCTCAGTTTCCAAAGTTCCAGGGGTTCCAGGTGTGCCCCCAA 204

## RESULT 15

BF320748/c  
LOCUS BF320748 470 bp mRNA EST 29-DEC-2000  
DEFINITION uz55hl1.y1 NCI-CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3673029 5' similar to SW:RFXA\_HUMAN 000287 REGULATORY FACTOR X-ASSOCIATED PROTEIN ; mRNA sequence.  
ACCESSION BF320748  
VERSION BF320748.1 GI:11269760  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 470)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Clone Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1433797

High quality sequence stop: 453.

## FEATURES

source Location/Qualifiers  
1. .470  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_lib="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 139 a 91 c 110 g 130 t

## ORIGIN

Query Match 2.3%; Score 53.4; DB 147; Length 470;  
Best Local Similarity 69.9%; Pred. No. 0.0046;  
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
Qy 1385 tctcttgagacagagtcctcactgtgtgcccaggtagtctcaaaacttgcggtccatt 1444  
Db 304 TCTCTTTTGAGACAGGAACCTTGCTATGTAGCCAGGCTGGCTTTGAACTTGGGATCCTCT 245  
Qy 1445 tgtctcactcatcagaatgctgggtccagggtgtgcacca 1487  
Db 244 TGGCTCAGTTTCCAAAGTTCCAGGGGTTCCAGGTGTGCCCCCAA 202

Search completed: June 21, 2001, 15:33:53

Job time: 7637 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:05:44 ; Search time 150.46 Seconds  
(without alignments)  
2798.729 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_8336\_10608

Perfect score: 2273

Sequence: 1 tctcgagacagacgcgtgt.....aggccttcctattgatcc 2273

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.6	2.1	7218	1	US-08-232-463-14
2	45.6	2.0	51259	3	US-08-781-891-209
3	43.2	1.9	152331	4	US-09-128-155-16
4	42.2	1.9	14507	3	US-08-785-150-1
5	40.8	1.8	152331	4	US-09-128-155-16
6	40.4	1.8	176373	4	US-08-128-155-17
7	40.2	1.8	380	1	US-08-126-587C-5
8	40	1.8	1920	1	US-08-087-772A-1
9	39.6	1.7	246240	2	US-08-724-394A-20
10	39.6	1.7	246240	2	US-08-724-394A-21
11	39.6	1.7	246240	2	US-08-724-394A-22
12	39.4	1.7	2880	4	US-09-115-954-3
13	39.4	1.7	3842	4	US-09-115-954-7
14	39.4	1.7	3912	4	US-09-115-954-1
15	39.2	1.7	11725	2	US-08-756-506-1
16	39.2	1.7	17041	1	US-08-076-011-1
17	39	1.7	35060	3	US-08-814-095-7
18	38.6	1.7	2908	3	US-08-487-799-1
19	38.6	1.7	6769	1	US-08-480-784-20
20	38.6	1.7	6769	1	US-08-483-553-20
21	38.6	1.7	6769	1	US-08-487-002-20
22	38.6	1.7	6769	1	US-08-483-554B-20
23	38.6	1.7	6769	4	US-08-488-011B-20
24	38.6	1.7	6769	4	US-08-850-727-20
25	38.6	1.7	6769	5	PCT-US95-10202-20
26	38.6	1.7	6769	5	PCT-US95-10203-20
27	38.6	1.7	6769	5	PCT-US95-10220-20

c	28	38.4	1.7	501	3	US-08-699-628-1	Sequence 1, Appl
	29	38.4	1.7	2520	2	US-08-454-557C-50	Sequence 50, Appl
	30	38.4	1.7	2520	2	US-08-340-426D-50	Sequence 50, Appl
	31	38.4	1.7	2520	2	US-08-450-673C-50	Sequence 50, Appl
	32	38.4	1.7	2520	5	PCT-US95-17111A-50	Sequence 50, Appl
c	33	38.2	1.7	5095	1	US-08-092-817-3	Sequence 3, Appl
	34	38	1.7	31571	1	US-08-323-443B-1	Sequence 1, Appl
c	35	38	1.7	40352	3	US-08-846-111D-15	Sequence 15, Appl
	36	38	1.7	53526	3	US-08-658-136-2	Sequence 2, Appl
	37	38	1.7	53577	3	US-08-658-136-1	Sequence 1, Appl
	38	37.6	1.7	615	4	US-08-998-416-186	Sequence 186, App
	39	37.4	1.6	7720	4	US-09-318-448-5	Sequence 5, Appl
	40	37	1.6	4084	3	US-08-866-340-1	Sequence 1, Appl
	41	37	1.6	4460	4	US-09-103-875-4	Sequence 4, Appl
	42	36.8	1.6	4922	2	US-08-330-272-5	Sequence 5, Appl
	43	36.8	1.6	4922	5	PCT-US95-13663-5	Sequence 5, Appl
c	44	36.6	1.6	5375	4	US-08-757-223-7	Sequence 7, Appl
c	45	36.6	1.6	10825	3	US-08-652-265-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

```

; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-781-891-209

Query Match                2.0%; Score 45.6; DB 3; Length 51259;
Best Local Similarity      62.1%; Pred. No. 0.0095;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1292  tattaatagttattctctctcttgagacagagctcactgtgtggccaggctagtct 1351
          ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23743  TTITAGTTTATGTTTACTTTTITTAGACAGGGCTCTCAGCTGTAGTGGGGACAGCT 23802

QY 1352  caaactgcggtccattgtctctcactcatcagaatgctgggctccagggtgtgc 1407
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23803  CCACCCCTGTCCCTTTTCTCTCACCCTCTGAGTGTCTGGGATCACAGCGCTGTGC 23858

RESULT 3
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-16

```

```

; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-153-16

Query Match      1.9%; Score 43.2; DB 4; Length 152331;
Best Local Similarity 63.5%; Pred. No. 0.095;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY   1261 tgcgaattagaaagcgtggtgctgcctacttataatagtattctttcttgaga 1320
      |||  || | ||||| | | | | | | | | | | | | | | | | | | | | |
Db   142531 tgcgatctacaggcggtgttttttttttttttttttttttttttcgtagaga 142590
      |||  || | ||||| | | | | | | | | | | | | | | | | | | | | |

QY   1321 cagagctcaactgtggccaggctagctcaaacctgcggtc 1364
      | | ||||| |||| | ||||| ||||| ||||| ||||| ||||| |||||
Db   142591 ctgggtctctctgtattgccaggctagctcgaacctcgggc 142634
      |||  || | ||||| | | | | | | | | | | | | | | | | | | | | |

RESULT          4
US-08-785-150-1
; Sequence 1, Application US/08785150
; Patent No. 6027915
; GENERAL INFORMATION:
; APPLICANT: Morris, Arvia E.
; APPLICANT: Lee, Chi-Chang
; APPLICANT: Thomas, James N.
; TITLE OF INVENTION: Expression Augmenting Sequence Elements
; Patent No. 6027915

```



```

ADDRESS: Bell, Seltzer, Park & Gibson
STREET: Post Office Drawer 34009
CITY: Charlotte
STATE: No. 5691155th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/087,772A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Linker, Raymond O.
REGISTRATION NUMBER: 26,419
REFERENCE/DOCKET NUMBER: 3339-195
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-881-3140
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 568..1731
US-08-087-772A-1

Query Match 1.8%; Score 40; DB 1; Length 1920;
Best Local Similarity 63.5%; Pred. No. 0.061;
Matches 61; Conservative 0; Mismatches 35; Indels 0

QY 1306 tctctctctgagacagagtctcactgtgtggtgccaggtagctctcaactgtgc
|||||
DB 97 TCTATATGTTGAGATAGTCTTTCTATATGTTGCCAGATTGCTGGACCTCA
|||||

QY 1366 attgtctcactcatcagaatgctgggcttcagggt 1401
|||||
DB 37 TTCTGCTCAACCTGCCAGTGTGGATTACAGAT 2
|||||

```

```

RESULT          9
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237rel
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 15:33:53 ; Search time 4579.66 Seconds  
(without alignments)  
4691.688 Million cell updates/sec

Title: US-09-445-201-1-COPY\_8336\_10608  
Perfect score: 2273  
Sequence: 1 tctcagacagacgcgtgt.....aggcttctctattgatcc 2273

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Capext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_estl1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*  
48: gb\_est48.\*  
49: gb\_est49.\*  
50: gb\_est50.\*  
51: gb\_est51.\*  
52: gb\_est52.\*  
53: gb\_est53.\*  
54: gb\_est54.\*  
55: gb\_est55.\*  
56: gb\_est56.\*  
57: gb\_est57.\*  
58: gb\_est58.\*  
59: gb\_est59.\*  
60: gb\_est60.\*  
61: gb\_est61.\*  
62: gb\_est62.\*  
63: gb\_est63.\*  
64: gb\_est64.\*  
65: gb\_est65.\*  
66: gb\_est66.\*  
67: gb\_est67.\*  
68: gb\_est68.\*  
69: gb\_est69.\*  
70: gb\_est70.\*  
71: gb\_est71.\*  
72: gb\_est72.\*  
73: gb\_est73.\*  
74: gb\_est74.\*  
75: gb\_est75.\*  
76: gb\_est76.\*  
77: gb\_est77.\*  
78: gb\_est78.\*  
79: gb\_est79.\*  
80: gb\_est80.\*  
81: gb\_est81.\*  
82: gb\_est82.\*  
83: gb\_est83.\*  
84: gb\_est84.\*  
85: gb\_est85.\*  
86: gb\_est86.\*  
87: gb\_est87.\*  
88: gb\_est88.\*  
89: gb\_est89.\*  
90: gb\_est90.\*  
91: gb\_est91.\*  
92: gb\_est92.\*  
93: gb\_est93.\*  
94: gb\_est94.\*  
95: gb\_est95.\*  
96: gb\_est96.\*  
97: gb\_est97.\*  
98: gb\_est98.\*  
99: gb\_est99.\*  
100: gb\_est100.\*  
101: gb\_est101.\*  
102: gb\_est102.\*  
103: gb\_est103.\*  
104: gb\_est104.\*  
105: gb\_est105.\*  
106: gb\_est106.\*  
107: gb\_est107.\*  
108: gb\_est108.\*  
109: gb\_est109.\*  
110: gb\_est110.\*  
111: gb\_est111.\*  
112: gb\_est112.\*  
113: gb\_est113.\*  
114: gb\_est114.\*  
115: gb\_est115.\*  
116: gb\_est116.\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	707	31.1	746	250	AZ815965	AZ815965 2M0084D06
2	560.8	24.7	586	242	AZ364007	AZ364007 1M0110F03
3	79.2	3.5	404	224	AQ126020	AQ126020 HS_3032.B
4	78.6	3.5	464	227	AQ332669	AQ332669 HS_5004.A
5	58.2	2.6	556	245	AZ508851	AZ508851 1M0351G04
6	56.6	2.5	506	248	AZ709930	AZ709930 RPCI-24-6
7	55	2.4	256	138	BE651040	BE651040 UI-M-BH3-
8	55	2.4	257	116	AW490147	AW490147 UI-M-BH3-
9	55	2.4	652	117	AW551517	AW551517 L0080A10-
10	54.8	2.4	559	243	AZ423401	AZ423401 1M0202C02
11	53.6	2.4	432	148	BF391061	BF391061 UI-R-CAL-
12	53.4	2.3	320	102	AI838610	AI838610 UI-M-A00-
13	53.4	2.3	429	23	AI661049	AI661049 mz78f09.x
14	53.4	2.3	431	102	AI849973	AI849973 UI-M-BG0-
15	53.4	2.3	470	147	BF320748	BF320748 uz55h11.y
16	53.4	2.3	478	2	AA119334	AA119334 mp80a10.r
17	53.4	2.3	631	121	AW826111	AW826111 us21h04.x
18	53	2.3	369	241	AZ306979	AZ306979 1M0008H17
19	52.4	2.3	311	109	AV102206	AV102206 AV102206
20	52.4	2.3	652	243	AZ403474	AZ403474 1M0171B14
21	52.2	2.3	621	251	AZ913824	AZ913824 RPCI-24-1
22	52	2.3	419	247	AZ637662	AZ637662 1M0497J12
23	51.8	2.3	395	109	AV096924	AV096924 AV096924
24	51.6	2.3	458	227	AQ355903	AQ355903 C1B1-E1-
25	51.2	2.3	584	248	AZ750042	AZ750042 RPCI-24-8
26	51.2	2.3	621	248	AZ744458	AZ744458 RPCI-24-1
27	51.2	2.3	631	243	AZ410081	AZ410081 1M0182G10
28	51.2	2.3	713	173	BC076348	BC076348 H3158F08-
29	50.8	2.2	446	17	AI171758	AI171758 EST217740
30	50.6	2.2	609	249	AZ791331	AZ791331 2M0041010
31	50.4	2.2	311	247	AZ662311	AZ662311 1M0541N17
32	50.2	2.2	367	138	BE655196	BE655196 UI-M-BG1-
33	50.2	2.2	376	241	AZ318225	AZ318225 1M0037H07
34	50.2	2.2	384	4	AA261578	AA261578 mz87a12.r
35	50.2	2.2	505	138	BE689422	BE689422 uw54d12.y
36	50.2	2.2	568	148	BF401678	BF401678 UI-R-CA0-
37	50.2	2.2	578	249	AZ762191	AZ762191 1M0556123
38	50.2	2.2	632	144	BF121097	BF121097 601757929
39	50.2	2.2	651	244	AZ461771	AZ461771 1M0267H18
40	50	2.2	489	1	AA004129	AA004129 mz82f03.r
41	50	2.2	631	243	AZ431614	AZ431614 1M0216G07
42	49.8	2.2	396	15	AI044824	AI044824 UI-R-C1-K
43	49.6	2.2	575	243	AZ393030	AZ393030 1M0155J19
44	49.6	2.2	624	246	AZ615574	AZ615574 1M0445D12
45	49.4	2.2	264	238	AZ100588	AZ100588 RPCI-23-2

## ALIGNMENTS

RESULT	1
AZ815965	
LOCUS	2M0084D06R Mouse 10kb DNA GSS 20-FEB-2001
DEFINITION	clone UUGC2M0084D06 R, DNA sequence.
ACCESSION	AZ815965
VERSION	AZ815965.1 GI:12985873
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 746) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE  
JOURNAL  
COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dduan@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0084 row: D column: 06  
Seq primer: CACACAGGAACACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 746.

FEATURES  
source

1. 746  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0084D06"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114[gb]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 162 a 159 c 169 g 256 t

## ORIGIN

Query Match 31.1% Score 707; DB 250; Length 746;  
Best Local Similarity 98.4%; Pred. No. 1.le-172;  
Matches 735; Conservative 0; Mismatches 10; Indels 2; Gaps 2;  
QY 1007 acacatgacatctattaaagttgggtatattcgggtggcatagctagctagctattgca 1066  
|||||  
Db 1 ACACATGATACATCTATTAGTGGGTATATTTCAGGGTGGCATAGCTAGACTATTGCA 60  
QY 1067 catctgtgtggtgagcagtgagaaactgctctgctgtctctcagaagccacagt 1126  
|||||  
Db 61 CATCTGTGTAGTGGCAGCTGGACAGCTGGCT-CTGGCTGTTCTCAGAGGCCACAGTG 119  
QY 1127 tcacggcatggctatttgccttggctctcttggcttaataacttattgacatggcctcatct 1186  
|||||  
Db 120 TCACGGCATTTGGCTATTGGCTTGGCTCTGCTTAATAACTTTATTGACATGGCCTCATCT 179  
QY 1187 tggttcagttcaactattttgcccacaaacacatcaatgccagctgagggccttagagtcac 1246  
|||||  
Db 180 TCGTTCAGTTCAGTTCATTTTCCCAACACAGCTCAATGCCAGCTAGAGGCTTAGAGTCA 239  
QY 1247 ctgtctcttagtcagtcgaattagaagcctggatgctgctgctctattatagttatt 1306  
|||||  
Db 240 CTGTTCTTAGTCAGTGTGGAATAGAAACCCCTGGATGCCCTGCTGCTATTATTAGTATT 299  
QY 1307 ctctctcttgagacagagtcactgctgctggtggtggccaggtgctgctcaacttgcggtcca 1366

```

|||||
Db 300 CTTCTCTTGAGACAGAGTCTCACTGTGTGGCCAGGCTAGTCTCAAACTTCGGGTCCA 359
QY 1367 ttgtctcactcatcagaatgctgggtccaggtgtgtgacacacactaggttagctcgc 1426
Db 360 TTTGTCTCACTCATCAGAATGCTGGGCTTCAGGTTGTGTCACACACTAGGTAGTCTCGC 419
QY 1427 gttttaagctaaagctggaagatcctgctgcttcccttaccatggtggcgatgttacaggt 1486
Db 420 GTTTTAAAGCTAAGAGCTGGAAGATCTGATGTCCTTTACCATGGTGGGCATGTTACAGGT 479
QY 1487 tagttgactgaaactagtattctcgtgtgtgaatgacctgcagtgatgtatctctca 1546
Db 480 TAGTTGACTGAAACTAGTTATCTCGCTGTGTAATGACCTGCAGTGGTATGATCTCTCA 539
QY 1547 agatgctttttgcatcttcaatcagttaggttaacagt-tcttaagcttccagcttgga 1605
Db 540 AGATGCTTTTGGCATTTCAATCAGTCTAGGTAAAGTCTCTTATGTCCTCAGCTTTGTA 599
QY 1606 ttggcatgagctcagagctttgatttaataagttgggacccccctagctattgtctattaga 1665
Db 600 TTGGTATGAGCTCAGAGCTTTGATTAAATAGTTGGGACCCCCCTAGCTATTGCTCATTAGA 659
QY 1666 ctacacatttttagtttctgctgagtttatgaatatgcatgcatgcatgcatgcatgcatg 1725
Db 660 CTTACACTATTTTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719
QY 1726 agatatttttctcccaattcctttt 1752
Db 720 AGATATTCTCTCTCCCAATTCCTTTT 746

RESULT 2
AZ364007 586 bp DNA GSS 02-OCT-2000
LOCUS 1M0110F03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0110F03 F, DNA sequence.
ACCESSION AZ364007
VERSION AZ364007.1 GI:10477707
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 586)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
-84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0110 row: F column: 03
Seq primer: CGTTGTAAGACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 586.
Location/Qualifiers
1..586
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0110F03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
FEATURES
source

```

```

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid p1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 137 a 125 c 137 g 186 t 1 others
ORIGIN

```

```

Query Match 24.7%; Score 560.8; DB 242; Length 586;
Best Local Similarity 98.5%; Pred. No. 9.7e-135;
Matches 576; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1338 gccagggttagtctcaaaactgctggtccattgtctcactcatcagatgctggctcc 1397
Db 2 GGCCTCAGCTAGTCTCAAACTTGGGCTCCATTGTCTCACTCATCAGAACTGCTGGCTTCC 61

QY 1398 aggtgtgtacacacactaggtgctgctgttttaagctaaagctggaagatcctgatg 1457
Db 62 AGGTGTGTGCACACACTAGGTAGTACGCTGCGCTTTTAAAGCTAAGAGCTGGAAGATCCTGATG 121

QY 1458 tctttaccatggtggcgtgttacaggttagttgactgaaactagttatcctgctgtg 1517
Db 122 TCTTTACCATGCTGGGGCATGTTACAGGTTAGTTGACTGAAACTAGTTATCTCGCTGTG 181

QY 1518 taatgacctgagtggtgtatctctcaagatgctttttgcatctcaatcagttaggt 1577
Db 182 TAATGACCTGCAGTGGTATGTATCTCTCAAGATGCTTTTGTGATTTCAATCAGTTAGGT 241

QY 1578 aacaagt-tcttaagctcctcagctgtgtatggcagctgagctgagctttgataatgag 1636
Db 242 ACAAAGTGTCTTATGTCTCCAGCTTTGTATGTTGATGAGCTCAGAGCTTTGATTAAATGAG 301

QY 1637 ttggaccctcctgctgcttattgctcattagcttacactatttttagtttgcctgagttt 1696
Db 302 TTGGACCCCTAGCTATTGCTCATAGACTTACACTATTTTGTGCTCTGAGTTT 361

QY 1697 atgaatgcatgtatgcatgaactgggagatatttttctcccaattccttttctc 1756
Db 362 ATGAATATGCATGTATGCATGAATGAGCTGAGAGATATTCTCTCTCCCAATTCCTTTTCC 421

QY 1757 cattaaatgctgtcttttagaagccactgctcagctcttgcagctcagataccacaaag 1816
Db 422 CATTTAAATGCTCTCTTTAGAAAGCCACTGCCTCAGCTTCTGCAGCTCAGATACCAAG 481

QY 1817 gaagtctgtcacagcatgataaaagacaattggacggggtcacagtggtcccgctccc 1876
Db 482 GAAGTCTGTGTACACAGCATGATAAAGACATGGGACGGGTGACAGTGGCTCCCGCTCCC 541

QY 1877 tttcaggggtatgagacgagctgtagagatgtctccagggag 1921
Db 542 TTTCAGGGGTATGGAGACGAGCTGTAGAGAGATGTCTCCAGGGAG 586

```

```

RESULT 3
AQ126020/c
LOCUS
DEFINITION

```

```

AQ126020 404 bp DNA GSS 23-SEP-1998
HS_3032_P2_C02_MF_C1T Approved Human Genomic Sperm Library D Homo

```

1 (bases 1 to 454)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
On Feb 19, 1999 this sequence version replaced gi:4130296.  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 580 row: A column: 17  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 464.  
Location/Qualifiers  
1..464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate:580 Col=17 Row=A"  
/clone\_lib="RPC111 Human Male BAC Library"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; RPC111 Human Male BAC Library"  
BASE COUNT 178 a 75 c 93 g 113 t 5 others  
ORIGIN  
Query Match 3.5%; Score 78.6; DB 227; Length 464;  
Best Local Similarity 66.5%; Pred. No. 1.4e-09;  
Matches 125; Conservative 0; Mismatches 62; Indels 1; Gaps 1;  
QY 1120 cacagtgcacgcatggctattgccttggctcttctgtactattat-tgacatgg 1178  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 298 CACTGGATGNTGCACGTGGGTATTTACTTATACTCTCTCTATTATTTCGGCTCATACTT 239  
QY 1179 cctcatctctgttccagcttcactatttgcaccaacacgcaatgccagctgagggcctta 1238  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 238 CGTCCTATTGTTTCATACTCTCTTATTGGCCACAGGTCAATGCCAGTTTAGGCCTAG 179  
QY 1239 ggagtcacatctgttcttagtcagtcgcgaattagaagcctggatgcctgcctgtattaat 1298  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 178 GGAGTCATTTTCTTAGTTGATGATGACTTAGAAGGCTTGGTGCTGCGCCACATCAAT 119  
QY 1299 taqgtatt 1306  
|| |||  
Db 118 TACTTTNT 111  
RESULT 5  
AZ508851/c 556 bp DNA GSS 05-OCT-2000  
LOCUS  
DEFINITION  
1M0351G04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0351G04 R, DNA sequence.  
ACCESSION  
AZ508851  
VERSION  
AZ508851.1 GI:10690167  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 556)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D. Weiss,R.





[5'-TGACGATGCTTAGATCGGAGCGCGCCCTTTTTTTTTTTT-3'] from 3.42 µg of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from







available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 1-41, >AT-rich#Low complexity 100-201, >PB1D10#SINE/Alu  
Seq primer: M13 Forward  
POLYA=Yes.

#### FEATURES

Location/Qualifiers  
1..431  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BG0-aib-b-06-0-UI"  
/clone\_lib="NIH\_BMAP\_MSC"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH\_BMAP\_MSC library is a non-normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806 by 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.  
TAG\_LIB="NIH\_BMAP\_MSC"  
TAG\_TISSUE="spinal-cord"  
TAG\_SEQ="TCGAA"

BASE COUNT 97 a 105 c 84 g 145 t

#### ORIGIN

Query Match 2.3%; Score 53.4; DB 102; Length 431;  
Best Local Similarity 69.9%; Pred. No. 0.0049;  
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 1309 tctctctgagacagagtctcactgtgtgcccaggtagtctcaaaccttgcgtccatt 1368  
Db 102 TCTCTTTTGAGACAGGAGCTGCTATGAGCCAGGCTGGCTTGAACCTGGGATCCTCT 161  
QY 1369 tatctcactcatcagaatcgtggtcaggtccagtggtgcacca 1411  
Db 162 TGGCTCAGTTCCTCAAGTTCCAGGGGTTCACAGGTGTCGCCCAA 204

#### RESULT 15

BF320748/c  
LOCUS BF320748 470 bp mRNA EST 29-DEC-2000  
DEFINITION uz55h11.y1 NCI-CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3673029 5' similar to SW:RFXA\_HUMAN 000287 REGULATORY FACTOR X-ASSOCIATED PROTEIN ; mRNA sequence.  
ACCESSION BF320748  
VERSION BF320748.1 GI:11269760  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 470)  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
TITLE Unpublished (1997)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov/image/html/iresources.shtml](http://image.llnl.gov/image/html/iresources.shtml)

MGI:1433797

High quality sequence stop: 453.

#### FEATURES

Location/Qualifiers  
1..470  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3673029"  
/clone\_lib="NCI-CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 139 a 91 c 110 g 130 t

#### ORIGIN

Query Match 2.3%; Score 53.4; DB 147; Length 470;  
Best Local Similarity 69.9%; Pred. No. 0.005;  
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 1309 tctctctgagacagagtctcactgtgtgcccaggtagtctcaaaccttgcgtccatt 1368  
Db 304 TCTCTTTTGAGACAGGAGCTGCTATGAGCCAGGCTGGCTTGAACCTGGGATCCTCT 245  
QY 1369 tgtctcactcatcagaatcgtggtcaggtccagtggtgcacca 1411  
Db 244 TGGCTCAGTTCCTCAAGTTCCAGGGGTTCACAGGTGTCGCCCAA 202

Search completed: June 21, 2001, 15:33:59

Job time: 7643 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:32:40 ; Search time 8015.43 Seconds  
(without alignments)  
993.820 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_10094\_10608  
Perfect score: 515  
Sequence: 1 tttaaatgtctgtctttag.....agggtcttctattggtatcc 515

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_hugo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_rod.\*  
33: em\_hum1.\*  
34: em\_hum2.\*  
35: em\_hum3.\*  
36: em\_hum4.\*  
37: em\_hum5.\*  
38: em\_hum6.\*  
39: em\_hum7.\*  
40: em\_in.\*  
41: em\_om.\*  
42: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vl2.\*  
60: gb\_hgl.\*  
61: gb\_hgl2.\*  
62: gb\_hgl3.\*  
63: gb\_hgl4.\*  
64: gb\_hgl5.\*  
65: gb\_hgl6.\*  
66: gb\_hgl7.\*  
67: gb\_hgl8.\*  
68: gb\_hgl9.\*  
69: gb\_hgl10.\*  
70: gb\_hgl11.\*  
71: gb\_hgl12.\*  
72: gb\_hgl13.\*  
73: gb\_hgl14.\*  
74: gb\_hgl15.\*  
75: gb\_hgl16.\*  
76: gb\_hgl17.\*  
77: gb\_hgl18.\*  
78: gb\_hgl19.\*  
79: gb\_hgl20.\*  
80: gb\_hgl21.\*  
81: gb\_hgl22.\*  
82: gb\_hgl23.\*  
83: gb\_hgl24.\*  
84: gb\_hgl25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rod.\*  
96: gb\_in4.\*  
97: gb\_pri0.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	515	100.0	12845	9	AX002124
2	510	99.0	510	94	AF061804
3	430	83.5	430	94	AF153058
C 4	304.6	59.1	135552	65	AC019251
C 5	304.6	59.1	214348	66	AC021220
C 6	254.4	49.4	126330	63	AC013745
C 7	39.4	7.7	847	13	AF327424
C 8	39.4	7.7	8785	12	AC007659

```

9 39.4 7.7 196646 80 AL365329 Mus muscu
10 39.4 7.7 259487 80 AL365322 Mus muscu
11 38.8 7.5 162299 75 AC079086 Homo sapi
12 38.8 7.5 194038 87 AC010103 Homo sapi
13 37 7.2 157721 83 AP003265 Oryza sat
14 37 7.2 182615 78 AF217246 Homo sapi
15 36.8 7.1 148267 81 AL390758 Homo sapi
16 36.4 7.1 99858 93 HSJ1059A9 Human DNA
17 36.4 7.1 187584 79 AL162731 Homo sapi
18 36.4 7.1 189352 62 AC011967 Homo sapi
19 36.4 7.1 202081 90 AL391987 Human DNA
20 36.4 7.1 215283 62 AC011966 Homo sapi
21 35.8 7.0 144075 63 AC013763 Homo sapi
22 35.6 6.9 176961 63 AC013356 Homo sapi
23 34.8 6.8 914 53 CNS01FJT Anopheles
24 34.8 6.8 141970 89 AL138761 Human DNA
25 34.8 6.8 149340 66 AC021514 Homo sapi
26 34.8 6.8 170909 80 AL360170 Homo sapi
27 34.6 6.7 32501 93 HSAP2B Homo sapi
28 34.6 6.7 133965 88 AF064861 Homo sapi
29 34.6 6.7 171648 77 AC087103 Papio ham
30 34.6 6.7 190349 87 AC009475 Homo sapi
31 34.6 6.7 191904 61 AC009956 Homo sapi
32 34.6 6.7 340000 92 HS21C079 Homo sapi
33 34.4 6.7 183575 82 AL590127 Homo sapi
34 34.4 6.7 184084 70 AC026659 Homo sapi
35 34.4 6.7 220242 64 AC016759 Homo sapi
36 34.2 6.6 4552 5 AF064258 Strangylo
37 34.2 6.6 125487 93 HSJ2685D6 Homo sapi
38 34.2 6.6 168074 75 AC079069 Homo sapi
39 34.2 6.6 185393 86 AC006552 Homo sapi
40 34 6.6 169845 65 AC013054 Homo sapi
41 34 6.6 175380 78 AC091045 Homo sapi
42 34 6.6 191466 73 AC068355 Homo sapi
43 33.8 6.6 140691 89 AF159056 Homo sapi
44 33.8 6.6 179137 82 AP000853 Homo sapi
45 33.8 6.6 217218 61 AC009132 Homo sapi

```

## ALIGNMENTS

```

RESULT 1
LOCUS AX002124 12845 bp DNA PAT 10-MAR-2000
DEFINITION Sequence 1 from Patent WO9855638.
ACCESSION AX002124
VERSION AX002124.1 GI:7241839
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 12845)
AUTHORS Breier,G. and Risau,W.
TITLE REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A
HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES
THEREOF
JOURNAL Patent: WO 9855638-A 1 10-DEC-1998;
MAX PLANCK GEFÄßSCHAFT (DE); BREIER GEORG (DE)
FEATURES
source 1..12845
/db_xref="taxon:32644"
BASE COUNT 3321 a 2893 c 2863 g 3723 t 45 others
ORIGIN
Query Match 100.0%; Score 515; DB 9; Length 12845;
Best Local Similarity 100.0%; Pred. No. 3.3e-154;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 tttaaatgctgtcttttagagcagcactccctcagcttctgcagctcagatacacaagga 60
|||||

```

```

Db 10094 TTAAATGCTGCTGCTTTAGAAAGCCACTGCCCTCAGCTTCTGCAGCTCAGATACCAAGGA 10153
Oy 61 agctgggtacacagcatgataaaacacaaatggaggggtcacagtggtcccgctccctt 120
|||||
Db 10154 AGTCTGGTACACAGCATGATAAAACACAAATGGACGGGTACAGTGGCTCCGTCCTT 10213
Oy 121 tcagggggtatggagacgagctgtagagagatgctccaggagggttttcattaatcagcaa 180
|||||
Db 10214 TCAGGGGTATGGACGACGAGCTGTAGAGAGATGCTCCAGGGAGTTTTCATTAAATCAGCAA 10273
Oy 181 tttaagtcagatctgtgacatctctcttaccagaataatgctgagtcgagtcagatca 240
|||||
Db 10274 TTTAGTCAGATCTGTGCATCTCTATGCTTTACAAGAAATGTCAGTGGGCTGAGATCATCA 10333
Oy 241 gatggaggttcacggttttcaatgctccgctatctcttttgaagacctgaaagtggcaa 300
|||||
Db 10334 GATGAGGTTTCATCGGTTTCAATGCTCCGATCTCTTTTGAAGACCTTGAAGTTGGCAA 10393
Oy 301 cgcgggaaaaacagaaactccacctcggcgccgtaattgcagagctgtgtgtgtgttg 360
|||||
Db 10394 CGCAGGAAAACAGGAACCTCCACCTGGTGGCCGTGAATTCGACAGAGCTGTGTGTGTGTG 10453
Oy 361 tgaccatctgccattctctctgttatcacagagctgtgaaactttaactggactgggg 420
|||||
Db 10454 TGACATCTGCCATCTCTCTGTTATGACAGAGCTTGGAACCTTTAACTGGAGTGGGG 10513
Oy 421 caaagtcaatccacacctttatatacaatgaattgctgaagaggccttttaaaacttgagtg 480
|||||
Db 10514 CAAAGTCAATCCACCTTTATACAATGAATGCTGAAGAGGCTTTTAAACTTGGAGTG 10573
Oy 481 tgcaattgttatgaaagggtttctctctctctctctctctctctctctctctctctctct 515
|||||
Db 10574 TGCATTGTTTATGGAAGGCGCTTCTCTATTGGATCC 10608
|||||
RESULT 2
AF061804 510 bp DNA ROD 23-AUG-1999
LOCUS Mus musculus fetal liver kinase 1 (Flkl) gene, intronic enhancer.
DEFINITION AF061804
ACCESSION AF061804
VERSION AF061804.1 GI:5757676
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 510)
TITLE Identification of vascular endothelial growth factor (VEGF)
receptor-2 (flk-1) promoter/enhancer sequences sufficient for
angioblast and endothelial cell-specific transcription in
transgenic mice
JOURNAL Blood 93 (12), 4284-4292 (1999)
MEDLINE 99290783
PUBMED 10361126
REFERENCE 2 (bases 1 to 510)
AUTHORS Kappel,A., Ronicke,V., Damert,A., Flamme,I., Risau,W. and
Breier,G.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1998) Molekulare Zellbiologie, Max Planck
Institut fuer physiologische und klinische Forschung, Parkstrasse
1, Bad Nauheim 61231, Germany
FEATURES
source Location/Qualifiers
1..510
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
1..510
/gene="Flkl"
/note="fetal liver kinase 1"
1..510
/gene="Flkl"
/note="intronic"

```

/db\_xref="taxon:10090"

McPheeters, R., Meldrim

McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Nayl

Norman, C.H., O'Connor, T., O'Donnell, P., Olivar, T.M., Peterson, K.,  
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
 Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## COMMENT

Direct Submission  
 Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 5, 2000 this sequence version replaced gl:6649482.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project information

Center project name: L2005

Center clone name: 5\_E\_20

-----

\* NOTE: This record contains 153 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 752: contig of 752 bp in length  
 853 852: gap of 100 bp  
 853 1637: contig of 785 bp in length  
 1638 1737: gap of 100 bp  
 1738 2506: contig of 769 bp in length  
 2507 2606: gap of 100 bp  
 2607 3392: contig of 786 bp in length  
 3393 3492: gap of 100 bp  
 3493 4261: contig of 769 bp in length  
 4262 4361: gap of 100 bp  
 4362 5128: contig of 767 bp in length  
 5129 5228: gap of 100 bp  
 5229 6000: contig of 772 bp in length  
 6001 6100: gap of 100 bp  
 6101 6856: contig of 756 bp in length  
 6857 6956: gap of 100 bp  
 6957 7734: contig of 778 bp in length  
 7735 7834: gap of 100 bp  
 7835 8636: contig of 802 bp in length  
 8637 8736: gap of 100 bp  
 8737 9526: contig of 790 bp in length  
 9527 9626: gap of 100 bp  
 9627 10374: contig of 748 bp in length  
 10375 10474: gap of 100 bp  
 10475 11244: contig of 770 bp in length  
 11245 11344: gap of 100 bp  
 11345 12128: contig of 784 bp in length  
 12129 12228: gap of 100 bp  
 12229 13009: contig of 781 bp in length  
 13010 13109: gap of 100 bp  
 13110 13926: contig of 817 bp in length  
 13927 14026: gap of 100 bp  
 14027 14758: contig of 732 bp in length  
 14759 14858: gap of 100 bp  
 14859 15655: contig of 797 bp in length  
 15656 15755: gap of 100 bp  
 15756 16543: contig of 788 bp in length  
 16544 16643: gap of 100 bp  
 16644 17436: contig of 793 bp in length  
 17437 17536: gap of 100 bp

17537 18309: contig of 773 bp in length  
 18310 18409: gap of 100 bp  
 18410 19183: contig of 774 bp in length  
 19184 19283: gap of 100 bp  
 19284 20038: contig of 755 bp in length  
 20039 20138: gap of 100 bp  
 20139 20878: contig of 740 bp in length  
 20879 20978: gap of 100 bp  
 20979 21780: contig of 802 bp in length  
 21781 21880: gap of 100 bp  
 21881 22642: contig of 762 bp in length  
 22643 22742: gap of 100 bp  
 22743 23549: contig of 807 bp in length  
 23550 23649: gap of 100 bp  
 23650 24442: contig of 793 bp in length  
 24443 24542: gap of 100 bp  
 24543 25313: contig of 771 bp in length  
 25314 25413: gap of 100 bp  
 25414 26164: contig of 751 bp in length  
 26165 26264: gap of 100 bp  
 26265 27045: contig of 781 bp in length  
 27046 27145: gap of 100 bp  
 27146 27921: contig of 776 bp in length  
 27922 28021: gap of 100 bp  
 28022 28812: contig of 791 bp in length  
 28813 28912: gap of 100 bp  
 28913 29663: contig of 751 bp in length  
 29664 29763: gap of 100 bp  
 29764 30555: contig of 792 bp in length  
 30556 30655: gap of 100 bp  
 30656 31427: contig of 772 bp in length  
 31428 31527: gap of 100 bp  
 31528 32317: contig of 790 bp in length  
 32318 32417: gap of 100 bp  
 32418 33198: contig of 781 bp in length  
 33199 33298: gap of 100 bp  
 33299 34081: contig of 783 bp in length  
 34082 34181: gap of 100 bp  
 34182 34938: contig of 757 bp in length  
 34939 35038: gap of 100 bp  
 35039 35795: contig of 757 bp in length  
 35796 35895: gap of 100 bp  
 35896 36624: contig of 729 bp in length  
 36625 36724: gap of 100 bp  
 36725 37501: contig of 777 bp in length  
 37502 37601: gap of 100 bp  
 37602 38375: contig of 774 bp in length  
 38376 38475: gap of 100 bp  
 38476 39264: contig of 789 bp in length  
 39265 39364: gap of 100 bp  
 39365 40148: contig of 784 bp in length  
 40149 40248: gap of 100 bp  
 40249 41001: contig of 753 bp in length  
 41002 41101: gap of 100 bp  
 41102 41871: contig of 770 bp in length  
 41872 41971: gap of 100 bp  
 41972 42757: contig of 786 bp in length  
 42758 42857: gap of 100 bp  
 42858 43646: contig of 789 bp in length  
 43647 43746: gap of 100 bp  
 43747 44507: contig of 761 bp in length  
 44508 44607: gap of 100 bp  
 44608 45362: contig of 755 bp in length  
 45363 45462: gap of 100 bp  
 45463 46193: contig of 731 bp in length  
 46194 46293: gap of 100 bp  
 46294 47090: contig of 797 bp in length  
 47091 47190: gap of 100 bp  
 47191 47981: contig of 791 bp in length  
 47982 48081: gap of 100 bp  
 48082 48867: contig of 786 bp in length  
 48868 48967: gap of 100 bp  
 48968 49724: contig of 757 bp in length



```
* 49725 49824: gap of 100 bp
* 49825 50586: contig of 762 bp in length
* 50587 50686: gap of 100 bp
* 50687 51456: contig of 770 bp in length
* 51457 51556: gap of 100 bp
* 51557 52334: contig of 778 bp in length
* 52335 52434: gap of 100 bp
* 52435 53204: contig of 770 bp in length
* 53205 53304: gap of 100 bp
* 53305 54077: contig of 773 bp in length
* 54078 54177: gap of 100 bp
* 54178 54958: contig of 781 bp in length
* 54959 55058: gap of 100 bp
* 55059 55833: contig of 775 bp in length
* 55834 55933: gap of 100 bp
* 55934 56716: contig of 783 bp in length
* 56717 56816: gap of 100 bp
* 56817 57612: contig of 796 bp in length
* 57613 57712: gap of 100 bp
* 57713 58422: contig of 710 bp in length
* 58423 58522: gap of 100 bp
* 58523 59310: contig of 788 bp in length
* 59311 59410: gap of 100 bp
* 59411 60186: contig of 776 bp in length
* 60187 60286: gap of 100 bp
* 60287 61061: contig of 775 bp in length
* 61062 61161: gap of 100 bp
* 61162 61891: contig of 730 bp in length

Query Match      59.1%; Score 304.6; DB 65; Length 135552;
Best Local Similarity 79.0%; Pred. No. 2e-86;
Matches 418; Conservative 0; Mismatches 94; Indels 17; Gaps 4;

QY 1 tttaaatgtgctgtcttttagaagccactgctcagcttctgcagctcagatcacatacaaga 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63576 TTTTATGTACATCTTTATGACCCATATTTCAGCTGGCGAGTTGGTTACACGGGA 63517

QY 61 agtcgtgtacacagcatgataaaagaacaatgggacggggtcacagtggctcccgctccct 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63516 AGCGCACTAGAAAA-TTGATAAAGGAAATGAGACAAGGTTCATAGATTCTCTCACTCCCT 63458

QY 121 tcagggtatgagacagctgtagagatgtctccaggagtttctaatcaatcagcaa 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63457 TCA-GGGTAGTGTAGTGAACATATATAAATCCGTCTAAGTGGGATTCGTTAATCAGCAA 63399

QY 181 tttagtcagatctgtcactctatgtcttacaagaaatgcagtgggcc----- 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63398 TTTAGTCAATGTGTACATCTTATGTTCTATAGAATGTCAGTGGGTCTTTCCCAAGG 63339

QY 230 ---tgagatcatcagatgaggttcatcggtttcaatgtcccgatccttttgaagac 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63338 GACTGAGATCATCAGATGAAGGTTTCATTTGGTTTCAATGTCCTGATCCCTTTTGTAAAGAC 63279

QY 287 cttagaagtgtgcaacagcaggaagaaacaccccttggtccggtgaattcagagct 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63278 CTTGAAGTGTGGCAATCAGGAAACAGAACTCCACCCCTAGTCATGAAATTCAGAACT 63219

QY 347 gttgtgtgtgtgtgacatctgcccattcttcctgttatgacagagctgtgaaactt 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63218 GTTGTGTTGTTTATGACCATCTGCCCATCTCTCTGTTATGACACAGCTTGTGAACCTT 63159

QY 407 aactgggaactgggggcaaaagtcaa-tcccaccttttataatgaattgtgaagagcctt 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63158 TACTGAGAATGGTGAAGAAAGTAAATTTCCACGTTTTATATAATGAATTCCTGAAGAGCCTT 63099

QY 466 ttaaaacttggaagtgcattgtttatggaagggcttctcatttgatc 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63098 TTAAGTATAGATATGCAATGTTTATGGAAGGTGTTTCTTCTATTAGGTC 63050

RESULT 5
AC021220/c AC021220 214348 bp DNA HTG 15-NOV-2000
LOCUS
```

## DEFINITION

Homo sapiens chromosome 4 clone RP11-530I17, WORKING DRAFT SEQUENCE, 38 unordered pieces.

## ACCESSION

AC021220

## VERSION

AC021220.4 GI:11178158

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 214348)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

Waterston,R.H.

## TITLE

The sequence of Homo sapiens clone

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 214348)

## AUTHORS

Waterston,R.H.

## TITLE

Direct Submission

## JOURNAL

Submitted (15-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA

## COMMENT

On Nov 15, 2000 this sequence version replaced gi:9309528.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: HNH0530117

----- Summary Statistics -----

Sequencing vector: M13; 87%

Sequencing vector: plasmid; 13%

Chemistry: Dye-primer ET; 87% of reads

Chemistry: Dye-terminator Big Dye; 13% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 195065 bases at least Q40

Consensus quality: 206600 bases at least Q30

Consensus quality: 203756 bases at least Q20

Insert size: 161000; agarose-fp

Insert size: 210648; sum-of-contigs

Quality coverage: 4.96 in Q20 bases; agarose-fp

Quality coverage: 4.62 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 1584: contig of 1584 bp in length  
\* 1585 1684: gap of unknown length  
\* 1685 4194: contig of 2510 bp in length  
\* 4195 4294: gap of unknown length  
\* 4295 6424: contig of 2130 bp in length  
\* 6425 6524: gap of unknown length  
\* 6525 9844: contig of 3320 bp in length  
\* 9845 9944: gap of unknown length  
\* 9945 14101: contig of 4157 bp in length  
\* 14102 14201: gap of unknown length  
\* 14202 21279: contig of 7078 bp in length  
\* 21280 21379: gap of unknown length  
\* 21380 31346: contig of 9967 bp in length  
\* 31347 31447: gap of unknown length  
\* 31447 40526: contig of 9080 bp in length  
\* 40527 40626: gap of unknown length  
\* 40627 52862: contig of 12236 bp in length  
\* 52863 66090: contig of 13128 bp in length  
\* 66091 66190: gap of unknown length  
\* 66191 82659: contig of 16459 bp in length  
\* 82660 82759: gap of unknown length  
\* 82760 104797: contig of 22038 bp in length  
\* 104798 104897: gap of unknown length





* 57629	57728:	gap of	100 bp
* 57729	58523:	contig of 795 bp in length	
* 58524	58623:	gap of	100 bp
* 58624	59367:	contig of 764 bp in length	
* 59388	59487:	gap of	100 bp
 Query Match      49.4%    Score 254.4; DB 63; Length 126330;			
Best Local Similarity 66.4%; Pred.No. 2.e-70;			
Matches 314; Conservative 0; Mismatches 144; Indels 15; Gaps 2;			
QY	57	aggaaagtctgtcacacagcatgataaaaagaacaatggacgggtccacagtgcgcccgcgc 116	
Dn	88902	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88843	
QY	117	ccttcagggggatggagacagcgtgtagagagatgctccagggaatttcaattaatca 176	
Dn	88842	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 88783	
QY	177	gcaatttagtcagactctgtcactccatgctttacaagaatatgcagtgggccc----- 229	
Dn	88782	GCAATTAGTCAAATGTTACATCCTAATGTTCATAAGAATGTCAGTGCGGTCTTTCC 88723	
QY	230	- - - - - tsgagatcatcagatggaggttcocggtttccaattgccgttatcccttgtta 282	
Dn	88722	AAGSGAGTGAGATCATCAGATGAAGGTTCAATTGGTTCCAATGCCGTATCCTTTGTA 88663	
QY	283	agacctgaagtggcgaacgcaggaacgaagaaactccaccctgggtgcgtgaattgcag 342	
Dn	88662	AGACTTTGAAGTTGGCAATGCAGAAACAGGAACCTCACCCTAGCTCCATGAATTCGAG 88603	
QY	343	agcttggtgttggttgcacatctgccactctccctatctctgtttatcacagagcttgttaa 402	
Dn	88602	AACGTGTTGTGTGGTTTTATGACCATCTGCCCATCTTCTCTGTATGACACAGCTTGTA 88543	
QY	403	ctttaactggggaactggggcgaagtcga-tccccacctttataaatgaattgctgaagag 461	
Dn	88542	CTTTTACTGAGAATGTCGAAAAGTAAATTCCTAGTTTATACAATGAATTCGTAAGAG 88483	
QY	462	ccctttaaaaactggagtgctattgtttatggaagggcttcttcattggatc 514	
Dn	88482	CCITTTAAAGTATAGATGTGCATGTGTTATGGAAGGTGTTCTTATTAGCTC 88430	
 RESULT 7			
AF327424/c			
LOCUS			
DEFINITION			
AP327424 847 bp mRNA PLN 26-FEB-2001			
Arabidopsis thaliana unknown protein (Tl4P1.19/At2g45010) mRNA,			
partial cds.			
AF327424			
ACCESSION AF327424.1 GI:11935196			
VERSION FLI CDNA.			
KEYWORDS			
SOURCE thale cress.			
ORGANISM Arabidopsis thaliana			
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
Brassicales; Brassicaceae; Arabidopsi.			
1 (bases 1 to 847)			
REFERENCE			
AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J.J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Eckert,J.R. and Theologis.A.			
Full Length cDNA of gene Tl4P1.19/At2g45010 (GI:4895249)			
Unpublished			
TITLE			
JOURNAL			
REFERENCE			
2 (bases 1 to 847)			
AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J.J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Eckert,J.R. and Theologis.A.			

TITLE	COMMENT
Davis, R.W., Ecker, J.R. and Theologis, A. Direct Submission Submitted (11-DEC-2000) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA	Riken Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA': Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.
The Salk, Stanford, PCEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banth, J., Chung, M.K., Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.	Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

```

FEATURES             source
Location/Qualifiers
1..847
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAFL3-4-K9 (R09069)"
/note="This clone is in pBluescript vector as a XhoI/SstI
insert.
ecotype: Columbia"
1..>847
/gene="t14P1.19/At2g45010"
1..133
/gene="t14P1.19/At2g45010"
134..>847
/gene="t14P1.19/At2g45010"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AAG42014.1"
/db_xref="GI:11935197"
/translation="WADGNASSRYKVLKEQAPVEEDITPGELNPIIDVPOLNVRKH
ECMVLPEYEPSSDENWTTFGCAEDPESCETGLFCPCVLFGRNI EAVERIPWTK
PCVCHAVCGEGMALAAITAFISGYIDPQTTVVICGLFFAMWCMGIGISGLFQELQK
KYLKNAPCDHCWVHCCLCAALCOEHREMKNHLSTDEASTTMDPPPPVQEMNTEER
RDASSSSSSPSSAKSQHND"
222 a 194 c 202 g 229 t
BASE COUNT
ORIGIN

```

	Query Match	7.7%	Score 39.4	DB 13	Length 847
	Best Local Similarity	71.2%	Pred. No. 0.19		
	Matches 52	Conservative 0	Mismatches 21	Indels 0	Gaps 0
QY	187	cagatctgtgcacatcctatgottttacaagaatatcagttggcctgagatcatcagatgga	246		
Db	360	CAGCTCTCTGGATCTTCAGCACACAACCAAAATGCCGGTGGTCCAGTTTTCATCAGAGGA	301		
QY	247	ggttcacatcggggtt	259		
Db	300	GGTTCATAGGTTT	288		

RESULT	8
AC007659/c	DNA
LOCUS	87885 bp
DEFINITION	Arabidopsis thaliana chromosome II section 241 of 255 of the complete sequence. Sequence from clones T13E15, T14P1.
ACCSSION	AC007659 AE02093
VERSION	AC007659.2 GI:6598773
KEYWORDS	HGT.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE AUTHORS

REFERENCE  
AUTHORS

1 (bases 1 to 87885)  
Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Reddy, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koehler, M.F., Cronin, L.A., Shen, M., VanKien, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhagen, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.

**TITLE**

JOURNAL  
NATURE 402 (1999)  
MEDLINE  
PUBMED

REFERENCE 2 (bases 1 to 87885)

REFERENCE  
AUTHORS  
Lin, X.

AUTHORS	TITLE	JOURNAL
WILLIAMS, J. H.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
SMITH, R. A.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
JOHNSON, D. E.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
MILLER, L. B.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
DAVIS, M. C.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
GARCIA, F. J.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
HARRIS, K. L.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
ROBERTS, N. P.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
TURNER, S. W.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
VAN DER MEULEN, J. P.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.
ZIMMERMAN, H. T.	THE EFFECT OF THE RATE OF GROWTH ON THE METABOLISM OF THE RAT	BIOCHEMICAL J.

**JOURNAL**

**COMMENT**

on Dec 17, 1999 this sequence version replaced gi:4092933.  
The sequence and annotation of chromosome 2 were merged from those  
of the individual clones on this chromosome after removing  
overlaps. For detailed information, please see the TIGR web site  
(<http://www.tigr.org/tadb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/Genscan.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13U16, the ESSA group for sequencing clone F1304, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

• Address all correspondence to: [at@tiqr.org](mailto:at@tiqr.org).

**FEATURES**  
**SOURCE**

```
misc_feature
    complement(<1..2901)
    /note="Sequence from clone T13E15"
misc_feature
    complement(2902..>87885)
misc_feature
    /note="Sequence from clone T14P1"
mRNA
    2904..>3809
```

```

/genes="At2g44940"
2304. .>3809
/genes="At2g44940"
/notes="T14P1.26; pFAM domain (PF00847)"
2322. .3809
/genes="At2g44940"
/codon_start=1
/product="putative AP2 domain transcription factor"
/db_xref="GI:4895256"
/translation="MARQINIESSVQVTFISSAIPAVSSSSITASASLSSTPTSS
SSSSNTNFIEDNKRKASRLSSLSVSDDDGGGGRKRTKDNKPTTSS
VMRSGKVGYSIREPKRSRLWGLPTAEMARADHGGRKRTKDNKPTTSS
ELTPPTNSPKDIOQAASLAANWQDSVNDVSEVAIEVEAPSRVAVQLFSSDTS
TTTTTQOSYSEASCASTACTDKDSEEEKFLDPLDFTDENMMIRNDACFYSSW
QLCGDAGFRLEPPFLSE"
complement(2972..3067)
/rpt_family="(GGA)n"
6558. .6661
/rpt_family="(TAAAA)n"
complement(6771..6870)
/rpt_family="(TAAAA)n"
complement(join(<8230..8361,8487..8652,8747..8949,
9035..9172,9411..9497,9590..9662,9774..9968,10342..10418,
10632..10703,10786..10946,11037..11255,11310..11343))
/genes="At2g44950"
complement(<8230..>11343)
/genes="At2g44950"
/notes="T14P1.25"
complement(join(8230..8361,8487..8652,8747..8949,
9035..9172,9411..9497,9590..9662,9774..9968,10342..10418,
10632..10703,10786..10946,11037..11255,11310..11343))
/genes="At2g44950"
/notes="unknown protein"
/codon_start=1
/protein_id="AAD32840.1"
/translation="MALLEKQLVSVVKKLSIRKQVEKDSIVWKEIKNKNELGV
SKTSVATDSMASLSEIQKQDERMIRKTRIGNTSRERGREIEADMKALISFPE
EMMSRQUNNKETAGGTHSLRADVSGVLCRKEKEYAEALQSLRAYSQGLDIN
ATVCDLKNHEELKFLDMYKRESTDARDIAEAKERYAWAHVQSLSDLEONLEL
RYKAANEAVSQOOLAAAEATADLRQKMDCKRDVAKHSDILKSKHEHGYLSIEI
QTIGSYEDIVFQNOQLLQVTERDDYNIKLFLEGTSRQMDTLIDKIKIMKDIOQ
GAYASFLSKSSRSLEQLQFCDFQKLAEDIKYSVLENQKLRSLRSLIEGSSAI
SKLESHSVESQSLDGALELEIRFNRRRIEEMETAKKKVSKRLSLIEGSSAI
QKLRQSLSEFKTELKACANDRPKEVVITKCYHLFCNCPVQKLTGTRQKKCPTCSAF
GPNDKPIVI"
complement(join(<12607..12809,12908..13135,13255..13336,
13417..13578,13759..13875))
/genes="At2g44960"
complement(<12607..>13875)
/notes="T14P1.24; similar to hypothetical protein
GB: AAC31828"
complement(join(12607..12809,12908..13135,13255..13336,
13417..13578,13759..13875))
/genes="At2g44960"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAD32839.1"
/db_xref="GI:4895254"
/translation="MASTGEPDKRRHFSISPSERAAAVKQFFWPPSSDKLDTAV
LQFQNLKSLQKLAQVECSILEDKLSQKEKQLPVNSSLKTVHKSWEKQLTASVSC
SVRVSSSGSAFVKNKDGSSPAVKNDFTNRLLETGATSSSNITCSNOMEENGAVT
SSMOTQTLNVAATLELCIKDELPTVLTNRILGDLQGLALSELSELEISFPRGDL
DVLVFKLSLRELQSHRDADAKVRVDRKLRIGSYLVSSHSTIYC"
14535. .14617
/rpt_family="(TAAAA)n"
complement(join(<14627..15100,15283..15570,15651..15738,
15995..16123,16289..16535,16728..16836,17072..17122))
/genes="At2g44970"
complement(<14627..>17122)
/genes="At2g44970"

/notes="T14P1.23; identical to GB: AAC31827"
complement(join(14627..15100,15283..15570,15651..15738,
15995..16123,16289..16535,16728..16836,17072..17122))
/genes="At2g44970"
/notes="unknown protein"
/codon_start=1
/protein_id="AAD32838.1"
/db_xref="GI:4895253"
/translation="MGEAGSDGSEVTPLVGGGNDNAVAPQVFNLSLPALNEAASYI
TOANSYLSGLEYGCKDTISHPHELLRSTSGVDGNSPVSCISGPRPFS"SSASDA
NSPRESTEIIPQATNAIVTSNRLNLGISMFOGLIERARRVVRGSAIDIGLQRAPE
MPVEDGDFRKNKILEIDGNHGPLYFDTTKFKSMGLACHIAKIHSSSEVKNAEII
KEYTEELGSKNRKVLHLLGSKGGYDAAALSYWPELKDVLGVLGASPGGSGTIA
TDIREGOLGTVNLRKMEILISKVIQDQIALEDITYERKKEFLNPHLEPTEPTV
SFRTASISPAVLSLTVSHVAHELPLTQAAKLPPVMPGLGMAAACAQLQVRYGKGS
DGLVTCDDAEVPGSVVVRKRLDHAWVYSVNEVLEADAQVCEALLTLVQVEQ
ERQKLA"KID"
complement(join(<18068..18186,18313..18477,18558..18703,
18775..19400,19477..19602,19724..19924,20006..20113,
20198..20287,20397..20514,20614..20749,20875..21046,
21269..21400,21551..21684,21782..21871,22075..22297))
/genes="At2g44980"
complement(<18068..>22297)
/genes="At2g44980"
/notes="T14P1.22"
complement(join(18068..18186,18313..18477,18558..18703,
18775..19400,19477..19602,19724..19924,20006..20113,
20198..20287,20397..20514,20614..20749,20875..21046,
21269..21400,21551..21684,21782..21871,22075..22297))
/genes="At2g44980"
/codon_start=1
/product="SNF2 subfamily global transcription activator"
/protein_id="AAD32837.1"
/db_xref="GI:4895252"
/translation="MSKSSPPKVPSTTMEYRRLEAAAEIILEKAKFSNTPPDCE
EGVATLKPQVGVSLQKYLKGVNVVLELDMGLKTLQALISFLKRPGLPG
PFLVCLPSVTDGWNVSEINRTPNLEVLYRGDKYCLRDMRKSMDHGFLEFDVLIT
TYDALVDQDELQIPQWYAIIDEAQRKNPNVLYNLEQLFLPRLLITGPTON
NLTEALMHFCMPLVFTGLDQAFKTEGDLVDNDKETYKSLKFLTGAFMLRRT
KSLITESCNIPLVPLTELTMVPLVSLQKITYSTILRKELPGLLELSGSGNHTSQN
IVIOLRKASHPYLFPETPEPFEHGLVQASGLLVLDQLLKRHDGSHRVLLFSQ
MSTFLDILQDMELRYSYERLDGVSRAERFAALKNFSAKTERGLDSEVDGSAEVE
MISPRAGVGLNLYAADVIFYEQDWNFQVQKALQRAHRIQISVLSINLVTESV
EEVILRAERKLQSLHNVDNNEEKEDGDLRSLVGLQRFDPPEIHNESDNLKM
VEISLAQVVAIRQNVDPKKEFEINSDTLTGNTSSALSDSEASYSLSWEK
LKEARSKDEKIIILGNRKNLSEERNLRIEAAKKAEEKKLATWGAGHYQSLSVEEP

Query Match 7.7%; Score 39.4; DB 12; Length 87885;
Best Local Similarity 71.2%; Pred. No. 0.35;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 187 cagatctgtcatctatcttcaagaatagtcagtgagcgcgcagatcatcagatgga 246
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34498 CAGCTCTCTGGATCTTCAGCACACCAACCAAAATGCCGCTGCTCCAGITTTTCATCAGAAGGA 34439

Qy 247 ggttcacatcgagttt 259
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34438 GGTTCATAGGTTT 34426

RESULT 9
AL365329 196646 bp DNA HTG 06-MAR-2001
LOCUS Mus musculus chromosome 1 clone RP23-181N5, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL365329
VERSION AL365329.6 GI:11863425
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 196646)

```

Lovell, J.  
 Direct Submission  
 Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
 requests: clonerequest@sanger.ac.uk  
 On Dec 15, 2000 this sequence version replaced gi:11691743.  
 ----- Genome Center  
 Center: UK Medical Research Council  
 Center code: UK-MRC  
 Web site: http://mrcseq.har.mrc.ac.uk  
 Contact: mouseq@har.mrc.ac.uk  
 ----- Project Information  
 Center project name: bm181N5  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of bases  
 Consensus quality: 196518 bases at least Q40  
 Consensus quality: 196567 bases at least Q30  
 Consensus quality: 196587 bases at least Q20  
 Insert size: 196846; sum-of-contigs  
 Insert size: 194063; 4.3% error; agarose-fp  
 Quality coverage: 15,00x in Q20 bases; sum-of-contigs Quality  
 coverage: 15.20x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

FEATURES
source
    Location/Qualifiers
    1. .196646
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /chromosome="1"
    /clone="RP23-181N5"
    /clone_lib="RPC1-23"
    1. .196646
    /note="assembly fragment:00126"

BASE COUNT  53272 a 39346 c 40064 g 63964 t
ORIGIN
Query Match      7.7%; Score 39.4; DB 80; Length 196646;
Best Local Similarity 53.6%; Pred. No. 0.39;
Matches 82; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY  145 gagagatgtctccaggaggttttcattaatcagcaatttagtcagatctgtgcacccaat 204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107247 GAGATANGTTTAGAGGCATTAGATATTTTCAGTCATTTTACACACITGTGGTTACTTTT 107306

QY  205 gctttacaagaaatgtcagtgggcctgagatcatcagatggaggttcacgggtttccaat 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107307 TACCTCCAAGTATTACCTGTGATCCTTACATAACTGAGCAATTTCCAGCAGTCTTAATG 107366

QY  265 gtcccgtatcctttgttaagaccttggaagttgg 297
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107367 TTTCCTTTCTTTTCTTTTGTGTTCTTGAAGTAG 107399

```

RESULT	10
AL365322	
LOCUS	AL365322
DEFINITION	Mus musculus chromosome 1 clone RP23-282D4, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION	AL365322
VERSION	AL365322.10 GI:12743807
KEYWORDS	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus 1 (bases 1 to 259487) Tromans,A.

Direct Submission  
 Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
 requests: clonerequest@sanger.ac.uk  
 On Feb 10, 2001 this sequence version replaced gi:12666279.  
 -----  
 Genome Center  
 Center: UK Medical Research Council  
 Center code: UK-MRC  
 Web site: http://mrcseq.har.mrc.ac.uk  
 Contact: mouseq@har.mrc.ac.uk  
 ----- Project Information  
 Center project name: BM282D4  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 99% of reads  
 Chemistry: Dye-primer Big Dye; 0% of reads  
 Consensus quality: 259474 bases at least Q40  
 Consensus quality: 259487 bases at least Q30  
 Consensus quality: 259487 bases at least Q20  
 Insert size: 259487; sum-of-contigs  
 Insert size: 243892; 2.4% error; agarose-fp  
 Quality coverage: 15.92x in Q20 bases; sum-of-contigs Quality  
 coverage: 17.27x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```
FEATURES
source

misc_ff

BASE COUNT
ORIGIN
```

[illegible]

RESULT 11  
 AC079086  
 LOCUS  
 DEFINITION Homo sapiens chromosome 4 clone RP11-70M19 map 4, WORKING DRAFT  
 SEQUENCE, 29 unordered pieces.  
 ACCESSION AC079086  
 VERSION AC079086.2 GI:11225392  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 162299)  
 REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 AUTHORS



TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens chromosome 4, clone RP11-70M19  
Unpublished  
2 (bases 1 to 162299)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bada,F., Boguslavsky,L.,  
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Darelliano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
Fitzhugh,W., Gage,D., Galaan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Taylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,  
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (18-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 18, 2000 this sequence version replaced gi:9845128.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10622  
Center clone name: 70\_M\_19  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 149171 bases at least Q40  
Consensus quality: 155807 bases at least Q30  
Consensus quality: 158244 bases at least Q20  
Insert size: 163000; agarose-ff  
Quality coverage: 3.7 in Q20 bases; agarose-ff  
Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 1088: contig of 1088 bp in length
* 1089 1188: gap of 100 bp
* 1189 2239: contig of 1051 bp in length
* 2240 2339: gap of 100 bp
* 2340 3811: contig of 1472 bp in length
* 3812 3911: gap of 100 bp
* 3912 5158: contig of 1247 bp in length
* 5159 5258: gap of 100 bp
* 5259 6908: contig of 1650 bp in length
* 6909 7008: gap of 100 bp
* 7009 8568: contig of 1560 bp in length
* 8569 8668: gap of 100 bp
* 8669 10761: contig of 2093 bp in length
* 10762 10861: gap of 100 bp
* 10862 13026: contig of 2165 bp in length

```

```

* 13027 13126: gap of 100 bp
* 13127 15329: contig of 2203 bp in length
* 15330 15429: gap of 100 bp
* 15430 17434: contig of 2005 bp in length
* 17435 17534: gap of 100 bp
* 17535 19935: contig of 2401 bp in length
* 19936 20035: gap of 100 bp
* 20036 23899: contig of 3864 bp in length
* 23900 23999: gap of 100 bp
* 24000 27393: contig of 3394 bp in length
* 27394 27493: gap of 100 bp
* 27494 47629: contig of 20136 bp in length
* 47630 47729: gap of 100 bp
* 47730 51364: contig of 3635 bp in length
* 51365 51464: gap of 100 bp
* 51465 55977: contig of 4513 bp in length
* 55978 56077: gap of 100 bp
* 56078 60357: contig of 4280 bp in length
* 60358 60457: gap of 100 bp
* 60458 64814: contig of 4357 bp in length
* 64815 64914: gap of 100 bp
* 64915 70287: contig of 5373 bp in length
* 70288 70387: gap of 100 bp
* 70388 75397: contig of 5010 bp in length
* 75398 75497: gap of 100 bp
* 75498 81554: contig of 6057 bp in length
* 81555 81654: gap of 100 bp
* 81655 88042: contig of 6388 bp in length
* 88043 88142: gap of 100 bp
* 88143 95604: contig of 7462 bp in length
* 95605 95704: gap of 100 bp
* 95705 102851: contig of 7147 bp in length
* 102852 102951: gap of 100 bp
* 102952 114380: contig of 11429 bp in length
* 114381 114480: gap of 100 bp
* 114481 126822: contig of 12342 bp in length
* 126823 126922: gap of 100 bp
* 126923 141300: contig of 14378 bp in length
* 141301 141400: gap of 100 bp
* 141401 159634: contig of 18234 bp in length
* 159635 159734: gap of 100 bp
* 159735 162299: contig of 2565 bp in length.

```

FEATURES  
Source  
1..162299  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-70M19"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..1088  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
1189..2239  
/note="assembly\_fragment"  
2340..3811  
/note="assembly\_fragment"  
3912..5158  
/note="assembly\_fragment"  
5259..6908  
/note="assembly\_fragment"  
7009..8568  
/note="assembly\_fragment"  
8669..10761  
/note="assembly\_fragment"  
10862..13026  
/note="assembly\_fragment"  
13127..15329  
/note="assembly\_fragment"  
15430..17434  
/note="assembly\_fragment"  
17535..19935

misc\_feature  
1..1088  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
1189..2239  
/note="assembly\_fragment"  
2340..3811  
/note="assembly\_fragment"  
3912..5158  
/note="assembly\_fragment"  
5259..6908  
/note="assembly\_fragment"  
7009..8568  
/note="assembly\_fragment"  
8669..10761  
/note="assembly\_fragment"  
10862..13026  
/note="assembly\_fragment"  
13127..15329  
/note="assembly\_fragment"  
15430..17434  
/note="assembly\_fragment"  
17535..19935





2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

```

the accession number will be preserved.
Location/Qualifiers
1. .157721
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0490D09"
BASE COUNT 44929 a 33979 c 34269 g 44344 t 200 others
ORIGIN
Query Match 7.2%; Score 37; DB 83; Length 157721;
Best Local Similarity 56.0%; Pred. No. 2.3;
Matches 70; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 197 cactcatgctttacaagaatgtcagtgggcctgagatcatcagatggaggttcacgg 256
||||| ||||| ||||| ||||| ||||| |||||
Db 24115 CATCCAATCTTTAGAAAGACATCCAGAGGTTCTGAGAGCCACAGATCAATCTTCAGTAA 24174
||||| ||||| ||||| ||||| ||||| |||||
QY 257 gtttcaatgcccgatctcttttgaagacctggaattggcacaacgcagaaaaacaggaa 316
||||| ||||| ||||| ||||| ||||| |||||
Db 24175 CTATCTGTGATGATTAATGATCAATCACCCTTCAAAGTTTGGTGCCAGTAAAGAGTCA 24234
||||| ||||| ||||| ||||| ||||| |||||
QY 317 ctcca 321
|||||
Db 24235 CTCCA 24239

```

RESULT	14
AF217246/c	
LOCUS	
DEFINITION	05-SEP-2000
ACCESSION	DNA HTG
VERSION	Homo sapiens chromosome 8 clone CTD-2017M2 map 8g24, WORKING DRAFT
KEYWORDS	SEQUENCE, 2l unordered pieces.
SOURCE	AF217246
ORGANISM	AF217246.4 GI:9967152 HTG; HTGS_PHASE1; HTGS_DRAFT. human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 182615) Taudien,S., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K., Rump,A., Schilhabel,M.B., Schudy,A., Wen,G. and Rosenthal,A. Chromosome 8 genomic sequence Unpublished 2 (bases 1 to 182615) Taudien,S., Wen,G.P., Schilhabel,M., Menzel,U., Jahn,N., Baumgart,C. and Rosenthal,A. Direct Submission Submitted (16-DEC-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany 3 (bases 1 to 182615) Taudien,S., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A., Wen,G. and Rosenthal,A. Direct Submission Submitted (05-SEP-2000) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany. Sequence update by submitter On Sep 5, 2000 this sequence version replaced gi:8152189. * NOTE: This is a 'working draft' sequence. It currently
JOURNAL	
TITLE	
REMARK	
COMMENT	

\* consists of 21 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1	9231:	contig of 9231 bp in length
9232	9331:	gap of unknown length
9332	10427:	contig of 1096 bp in length
10428	10527:	gap of unknown length
10528	12883:	contig of 2356 bp in length
12884	12983:	gap of unknown length
12984	14772:	contig of 1789 bp in length
14773	14872:	gap of unknown length
14873	18731:	contig of 3859 bp in length
18732	18831:	gap of unknown length
18832	22448:	contig of 3617 bp in length
22449	22548:	gap of unknown length
22549	26063:	contig of 3515 bp in length
26064	26163:	gap of unknown length
26164	30412:	contig of 4249 bp in length
30413	30512:	gap of unknown length
30513	37955:	contig of 7343 bp in length
37956	37955:	gap of unknown length
37956	48596:	contig of 10641 bp in length
48597	48696:	gap of unknown length
48697	58215:	contig of 9519 bp in length
58216	58315:	gap of unknown length
58316	72799:	contig of 14484 bp in length
72800	72899:	gap of unknown length
72900	86382:	contig of 13483 bp in length
86383	86482:	gap of unknown length
86483	100353:	contig of 13871 bp in length
100354	100453:	gap of unknown length
100454	113432:	contig of 12979 bp in length
113433	113532:	gap of unknown length
113533	130872:	contig of 17340 bp in length
130873	130972:	gap of unknown length
130973	156022:	contig of 25050 bp in length
156023	156122:	gap of unknown length
156123	178374:	contig of 22252 bp in length
178375	178474:	gap of unknown length
178475	179488:	contig of 1014 bp in length
179489	179588:	gap of unknown length
179589	181002:	contig of 1414 bp in length
181003	181102:	gap of unknown length
181103	182615:	contig of 1513 bp in length
181104	182615:	contig of 1513 bp in length

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```
source
1. 102010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q24"
/clone="CTD-2017M2"
1. 9231
/misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
```

BASE COUNT	57804	a	35251	c	33689	g	53742	t	2129	others
ORIGIN	vector_side:left									

Query Match	7.2%	Score 37;	DB 78;	Length 182615;
Best Local Similarity	56.0%;	Pred. No. 2.3;		
Matches 70;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0;

[illegible]

Db 128518 TTGCTATTCAAGATGGGGTGAGAAAAAGAACAGCTGACCCCTTGACCAATACAGTTAGA 128459

QY 457 agagg 461

Db 128458 ACTGG 128454

# RESULT 15

AL390758/c

LOCUS

DEFINITION

PROGRESS \*\*\*, 21 unordered pieces.

ACCESSION

AL390758

VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Sims, S.

Direct Submission

Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SP, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 25, 2000 this sequence version replaced gi:9716790.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA540C21

----- Summary Statistics

Sequencing program: XGAP4; version 4.5

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 136465 bases at least Q40

Consensus quality: 141338 bases at least Q30

Consensus quality: 144184 bases at least Q20

Insert size: 146267; sum-of-contigs

Insert size: 166241; 2.3% error; agarose-fp

Quality coverage: 3.25x in Q20 bases; sum-of-contigs Quality

coverage: 3.06x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 21 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 4224: contig of 4224 bp in length

\* 4225 4324: gap of 100 bp

\* 4325 10808: contig of 6484 bp in length

\* 10809 10908: gap of 100 bp

\* 10909 26777: contig of 15869 bp in length

\* 26778 26877: gap of 100 bp

\* 26878 31189: contig of 4312 bp in length

\* 31190 31289: gap of 100 bp

\* 31290 40015: contig of 8726 bp in length

\* 40016 40115: gap of 100 bp

\* 40116 52786: contig of 12671 bp in length

\* 52787 52886: gap of 100 bp

\* 52887 55454: contig of 2568 bp in length

\* 55455 55554: gap of 100 bp

\* 55555 59896: contig of 4342 bp in length

\* 59897 59996: gap of 100 bp

\* 59997 64419: contig of 4423 bp in length

\* 64420 64519: gap of 100 bp

\* 64520 66621: contig of 2102 bp in length

\* 66622 66721: gap of 100 bp

\* 66722 70075: contig of 3354 bp in length

\* 70076 70175: gap of 100 bp

\* 70176 88176: contig of 18001 bp in length

\* 88177 88276: gap of 100 bp

\* 88277 90769: contig of 2493 bp in length

\* 90770 90869: gap of 100 bp

\* 90870 96615: contig of 5746 bp in length

\* 96616 96715: gap of 100 bp

\* 96716 101018: contig of 4303 bp in length

\* 101019 101118: gap of 100 bp

\* 101119 106502: contig of 5384 bp in length

\* 106503 106602: gap of 100 bp

\* 106603 109577: contig of 2975 bp in length

\* 109578 109677: gap of 100 bp

\* 109678 129521: contig of 19844 bp in length

\* 129522 129621: gap of 100 bp

\* 129622 134344: contig of 4723 bp in length

\* 134345 134444: gap of 100 bp

\* 134445 142809: contig of 8365 bp in length

\* 142810 142909: gap of 100 bp

\* 142910 148267: contig of 5358 bp in length.

## FEATURES

source

1. .148267

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10"

/clone="RP11-540C21"

/clone\_lib="RPC1-11.2"

1. .4224

/note="assembly\_fragment:00232"

fragment\_chain:1

4325. .10808

/note="assembly\_fragment:00914"

fragment\_chain:1

10909. .26777

/note="assembly\_fragment:00153"

26878. .31189

/note="assembly\_fragment:00185"

31290. .40015

/note="assembly\_fragment:00235"

40116. .52786

/note="assembly\_fragment:00323"

52887. .55454

/note="assembly\_fragment:00363"

55555. .59896

/note="assembly\_fragment:00401"

59997. .64419

/note="assembly\_fragment:00466"

64520. .66621

/note="assembly\_fragment:00473"

66722. .70075

/note="assembly\_fragment:00517"

70176. .88176

/note="assembly\_fragment:00531"

88277. .90769

/note="assembly\_fragment:00542"

90870. .96615

/note="assembly\_fragment:00568"

96716. .101018

/note="assembly\_fragment:00571"

101119. .106502

/note="assembly\_fragment:00784"

106603. .109577

/note="assembly\_fragment:00869"

109678. .129521

/note="assembly\_fragment:00907"

129622. .134344

/note="assembly\_fragment:00990"

134445. .142809

/note="assembly\_fragment:01164"

142910. .148267

/note="assembly\_fragment:00253"

clone\_end:SP6

Search completed: June 21, 2001, 18:37:34  
Job time: 18633 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:00:40 : Search time 346.98 Seconds  
(without alignments)  
866.469 Million cell updates/sec

Title: US-09-445-201-1-copy\_10094\_10608

Perfect score: 515

Sequence: 1 tttaaatgtctgtcttag.....agggttcctcttgatcc 515

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0401.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	515	100.0	12845	20 V84274	Mouse endothelial
C 2	39.4	7.7	1044	21 C45640	Arabidopsis thalia
C 3	39.4	7.7	1046	21 C33819	Arabidopsis thalia
C 4	33.8	6.6	2274	20 X84333	Stealth virus nucl
5	32.2	6.3	92334	21 A81473	N. meningitidis pa
6	32.2	6.3	349980	21 F21544	Neisseria mening
7	32.2	6.3	1437668	21 A81490	N. meningitidis B
C 8	31.6	6.1	5520	21 C51145	Arabidopsis thalia
9	30.8	6.0	1222	20 X37475	Human secreted pro
10	30.8	6.0	5612	21 C76389	Human ORFX ORF1944
11	30.6	5.9	314	19 V31227	E. coli J96 pathog

12	30.6	5.9	11165	21 A15186	DNA encoding Esche
13	30.6	5.9	273254	21 C81914	Chlamydia pneumoni
14	30.4	5.9	3161	21 C74253	Human secreted pro
C 15	30	5.8	692	20 Z42012	Human endometrium
16	30	5.8	7679	21 Z36322	Mechanical stress
17	30	5.8	8882	21 Z36321	Mechanical stress
18	29.8	5.8	550	21 C52708	Arabidopsis thalia
C 19	29.8	5.8	1590	21 Z99655	DNA coding for the
20	29.8	5.8	2105	21 C59480	Human secreted pro
21	29.8	5.8	2432	21 A79269	Pinus radiata cell
C 22	29.6	5.7	308	21 C16102	Human secreted pro
23	29.6	5.7	568	21 X06677	Human colon cancer
C 24	29.6	5.7	2693	20 X02976	Human IL-1ra BAC c
25	29.4	5.7	704	22 F44760	Annexin-like prote
26	29.4	5.7	2429	21 A98003	Human T gene DNA f
27	29.2	5.7	648	21 F10383	Fusarium venenatum
28	29	5.6	397	8 N71366	Sequence encoding
C 29	29	5.6	837	21 Z36301	Mechanical stress
30	29	5.6	886	12 Q10077	Engineered human T
31	29	5.6	886	17 T41697	TSH beta subunit g
C 32	29	5.6	1947	20 X52247	Protein PR0239 cDN
C 33	29	5.6	2384	19 V30919	Human secreted pro
C 34	29	5.6	2641	21 Z54227	HTM clone 2709055
C 35	29	5.6	5077	21 A50096	Arabidopsis herbic
36	28.8	5.6	535	21 C54997	Arabidopsis thalia
37	28.8	5.6	704	21 A67096	Pinus radiata anne
38	28.8	5.6	704	21 A67132	Pinus radiata anne
C 39	28.8	5.6	1420	20 X90222	GRK4 polymorphism
40	28.8	5.6	1815	21 Z54283	Neisseria gonorrhoe
41	28.8	5.6	2088	18 T96687	Human TUB form 6 c
42	28.8	5.6	2088	21 A94680	Human TUB form 6 c
43	28.6	5.6	440	21 F15627	Human prostate can
44	28.6	5.6	716	22 C90339	AI378857 cDNA clon
C 45	28.6	5.6	2346	21 A95821	Human metalloprote

#### ALIGNMENTS

RESULT 1  
V84274  
ID V84274 standard; DNA; 12845 BP.  
XX AC V84274;  
XX DT 12-APR-1999 (first entry)  
XX DE Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.  
XX KW Flk-1; endothelial growth factor receptor-2; VEGF; transcription;  
KW endothelium; enhancer; vascular disease; angiogenesis; cancer;  
KW diabetic retinopathy; rheumatoid arthritis; wound healing;  
KW vulneryary; atherosclerosis; tumour; neuronal disorder; therapy;  
KW diagnosis; mouse; ss.  
XX OS Mus sp.  
XX FH Key Location/Qualifiers  
FT promoter 6036..6959  
FT exon /tag= a  
FT 6661..7026  
FT /tag= b  
FT /number= 1  
FT /codon\_start= 6960  
FT intron 7027..10632  
FT /tag= c  
FT /number= 1  
FT /note= "DNA sequences comprising nucleotides  
8260-10560, 8336-10608 and/or 10094-10608  
are specifically claimed in Claim 3b"  
FT protein\_bind 10137..10155  
FT /tag= d  
FT /note= "GATA/PEA3 consensus binding site"

protein\_bind 10166..10179  
 /tag= e  
 /note= "GATA consensus binding site"  
 10187..10197  
 /tag= f  
 /note= "AP1 consensus binding site"  
 10252..10261  
 /tag= g  
 /note= "NFkB consensus binding site"  
 10273..10283  
 /tag= h  
 /note= "AP1 consensus binding site"  
 10301..10309  
 /tag= i  
 /note= "STAT consensus binding site"  
 10356..10372  
 /tag= j  
 /note= "Ets-1/GATA consensus binding site"  
 10395..10404  
 /tag= k  
 /note= "Ets-1 consensus binding site"  
 10453..10477  
 /tag= l  
 /note= "SCL/TAL-1 Ets-1 consensus binding site"  
 10643..10726  
 /tag= m  
 /number= 2  
 10727..12673  
 /tag= n  
 /number= 2  
 12674..12845  
 /tag= o  
 /number= 3

WO9855638-Al.  
 10-DEC-1998.  
 03-JUN-1998; 98WO-EF03318.  
 03-JUN-1997; 97EP-0108959.  
 (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 Breier G, Risau W, Roenicke V;  
 WPI; 1999-059915/05.

New recombinant DNA containing heterologous sequence controlled by endothelial cell specific regulator - from the Flk-1 gene, used to treat, prevent or diagnose vascular disease, tumours, also to screen for transcriptional regulators

Claim 3; Fig 1; 107pp; English.

This 12.8 kb DNA sequence spans the region from -6.65 kb relative to the transcriptional start site to +6.15 (located in the third exon) of the murine endothelial growth factor (VEGF) receptor-2 gene Flk-1. This portion of the Flk-1 gene was isolated from the liver of mouse strain 129/SvJ by screening a phage library of liver genomic DNA in vector lambda FixII. Sequences within the 5'-flanking region of the gene, in combination with sequences located within the first intron, specifically and reproducibly target expression of heterologous DNA to angioblasts during early stages of vascular development and also to the vasculature of postnatal mice. The regulatory sequences of the first intron also function as an autonomous endothelium-specific enhancer when fused to a heterologous promoter (e.g. HSV-TK promoter). This Flk-1 intron enhancer contains several potential binding sites for transcription factors of the Ets and GATA families. The invention provides a new claimed recombinant DNA (I) comprising at least one regulatory sequence from an intron of the Flk-1 gene, or its homologue, to control expression in endothelial cells, in vivo, linked to (b) a

heterologous DNA. (I) is used to produce transgenic animals, and these, or transformed cells, are used to identify agents (A), potential pharmaceuticals, that suppress, activate or enhance transcription of genes in endothelial cells. (I), vectors and (A) are used to direct or prevent (for antisense sequences) expression of genes specifically in endothelial cells, e.g. for treating angiogenesis, cancer, diabetic retinopathy, rheumatoid arthritis etc., and in wound healing, particularly for treating vascular disease (particularly proliferation of smooth muscle cells, specifically atherosclerosis), tumours and neuronal disorders. They may also be used to induce vascular disease in the transgenic animals or diagnostically, particularly for studying (mal)function, interactions and unregulated expression of endothelial products. (I) provide specific modulation of gene expression in endothelial cells, at all stages of development.

Sequence 12845 BP; 3321 A; 2893 C; 2863 G; 3723 T; 45 other;

Query Match 100.0%; Score 515; DB 20; Length 12845;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-164;  
 Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tttaaatgtgtgttttagaagccactgcctcagcttctgagctcagatccaaagga 60  
 Db 10094 tttaaatgtgtgttttagaagccactgcctcagcttctgagctcagatccaaagga 10153  
 Qy 61 agtctgttacacagcatgataaaagacaaatggaggggtcacagtggctcccgctccctt 120  
 Db 10154 agtctgttacacagcatgataaaagacaaatggaggggtcacagtggctcccgctccctt 10213  
 Qy 121 tcaggggtatggagacgagctgtagagagatgtctccagggaggttttcattaatcagcaa 180  
 Db 10214 tcaggggtatggagacgagctgtagagagatgtctccagggaggttttcattaatcagcaa 10273  
 Qy 181 tttagtcagatctgtgcattccttatctttacaagaatgtcagtgaggcctgagatcatca 240  
 Db 10274 tttagtcagatctgtgcattccttatctttacaagaatgtcagtgaggcctgagatcatca 10333  
 Qy 241 gatggaggttcacgtgggtttcaatgtcccgatcccttttgaagaccttgaagtggcaa 300  
 Db 10334 gatggaggttcacgtgggtttcaatgtcccgatcccttttgaagaccttgaagtggcaa 10393  
 Qy 301 cgcagggaaaacaggaactccaccctgggtgcttggaattgcagagctgtgtgtgtttg 360  
 Db 10394 cgcagggaaaacaggaactccaccctgggtgcttggaattgcagagctgtgtgtgtttg 10453  
 Qy 361 tgaccatctgccattcttcctgttatgacagagctgtgaaactttaactggactgggg 420  
 Db 10454 tgaccatctgccattcttcctgttatgacagagctgtgaaactttaactggactgggg 10513  
 Qy 421 caaagtcacatccaccctttatacaatgaattgtgaagagggccttttaaaacttggagt 480  
 Db 10514 caaagtcacatccaccctttatacaatgaattgtgaagagggccttttaaaacttggagt 10573

RESULT 2  
 C45640/c  
 ID C45640 standard; DNA; 1044 BP.  
 XX C45640;  
 AC C45640;  
 XX 18-OCF-2000 (first entry)  
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 47226.  
 DE Hybridisation assay; genetic mapping; gene expression control;  
 XX protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.





```
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.7%; Score 39.4; DB 21; Length 1046;
Best Local Similarity 71.2%; Pred. No. 0.0037;
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 187 cacatctgtcatcctatgctttacaagaatgtcaatggcctgagatcatcatcagatgga 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 CAGCTCTCTGGATCTTCAGCACACCAAAATGCCGTGGTCCAGTTTTCATCAGAAAGGA 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 247 ggttcacgcgggtt 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 GGTTCATAGGTTT 290

RESULT 4
X84333/c
ID X84333 standard; DNA; 2274 BP.
XX
AC X84333;
XX
XX
DT 08-SEP-1999 (first entry)
XX
DE Stealth virus nucleic acid clone, SEQ ID NO: 25.
XX
DE Stealth virus; detection; diagnosis; infection; ss.
XX
OS Stealth virus.
XX
FH Key Location/Qualifiers
FT misc_difference 2196
FT tag- a
FT note- "this nucleotide is represented as a * in the
FT specification, and is included to maintain the
```







```

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 6.1%; Score 31.6; DB 21; Length 5520;
Best Local Similarity 55.5%; Pred. No. 3.9;
Matches 61; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 67 gtacacagcatgataaaagacaaatggacgggtgcacagtggtccctcccttcagg 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3739 GCACATAAGATGAAAGAAAGGTGAGAGTGGTCACTTGTCTAATGCCCCGTTCTGGT 3680

QY 127 gtaggagagagctgtagagagatgtctccaggaggttttcattatca 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3679 TGTCAAACCAAGTCGAGAGAGAAAGCTTCAAGGAGATTCCATTCAACA 3630

RESULT 9
X37475
ID X37475 standard: cDNA; 1222 BP.
XX
AC X37475;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 25.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.
XX
PN WO9918208-A1.
XX
PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US20775.
XX
PR 02-OCT-1997; 97US-0060884.
PR 02-OCT-1997; 97US-0060833.
PR 02-OCT-1997; 97US-0060836.
PR 02-OCT-1997; 97US-0060837.
PR 02-OCT-1997; 97US-0060838.
PR 02-OCT-1997; 97US-0060839.
PR 02-OCT-1997; 97US-0060843.
PR 02-OCT-1997; 97US-0060862.
PR 02-OCT-1997; 97US-0060866.
PR 02-OCT-1997; 97US-0060874.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Duan DR, Endress GA, Feng P, Ferlie AM;
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
DR WPI; 1999-264022/22.
DR P-PSDB; Y07876.
XX
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1a; page 241-242; 368pp; English.
XX
CC This invention describes novel isolated human genes and the secreted

```









XX 19-MAR-1999; 99US-0125362.  
PR 10-DEC-1999; 99US-0169980.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen GA, Ruben SM, Komatsoulis G;  
XX WPI; 2000-579483/54.  
DR P-PSDB; B39209.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
PT used in preventing, treating or ameliorating a medical condition -  
XX  
XX Claim 1; Page 357-358; 434pp; English.  
XX  
XX The polynucleotide sequences given in C74223-C74279 encode the human  
CC secreted proteins represented in B39179-B39226. Sequences B39227-B39308  
CC are alternative proteins encoded by the genes, and also protein sequences  
CC with which they share homology. The proteins have activities based on the  
CC tissues and cells in which they are expressed. Examples of activities  
CC include: immunosuppressive; antiarthritic; antirheumatic;  
CC antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;  
CC antitropic; neuroprotective; antibacterial; virucide; fungicide; and  
CC ophthalmological. The human secreted proteins, polynucleotides,  
CC antagonists and agonists of the invention may be useful in the treatment,  
CC prevention, and/or diagnosis of various disease, disorders and conditions  
CC such as autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. The polypeptides can also be  
CC used to aid wound healing and epithelial cell proliferation, to  
CC regenerate tissues, maintain organs before transplantation, in chemotaxis  
CC and as a food additive or preservative e.g. to increase storage  
CC capabilities. Sequences C74214-C74222 and B39178 are used during the  
CC isolation and characterisation of the genes of the invention.  
XX  
SQ Sequence 3161 BP; 1025 A; 485 C; 603 G; 1048 T; 0 other;

Query Match 5.9%; Score 30.4; DB 21; Length 3161;  
Best Local Similarity 50.7%; Pred. No. 7.4;  
Matches 73; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 372 ccattcttctgttatgacagagctgtgaaactttaactgggactggggcaagtaac 431  
DB 1716 ccatttgtgaatcatcaagaagatgttaactctgcttttagcattgatgcaagttataa 1775  
QY 432 ccacctttatacaatgaattgtgaaaggcccttttaaaacttgagtgatgctgttta 491  
DB 1776 taactgacattaaaaaatacatttgggaagctttattgaatatattgagataagggttaa 1835  
QY 492 tgaaggggcttctcattgatcc 515  
DB 1836 tggagtcctctctcattgacac 1859

RESULT 15  
Z42012/C  
ID Z42012 standard; cDNA; 692 BP.  
XX  
XX Z42012;  
XX  
XX 31-JAN-2000 (first entry)  
DE Human endometrium tumour cDNA derived EST 32.  
XX  
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
KW treatment; uterine; gene therapy; expressed sequence tag; ss.  
XX  
OS Homo sapiens.

XX DE19817948-A1.  
PN 21-OCT-1999.  
XX  
XX 17-APR-1998; 98DE-1017948.  
PF  
XX  
XX 17-APR-1998; 98DE-1017948.  
PR  
XX  
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
PI WPI; 1999-591957/51.  
XX P-PSDB; Y60034, Y60035, Y60036.  
DR  
XX  
XX New nucleic acid sequences expressed in uterine cancer tissues, and  
PT derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents -  
XX  
XX Claim 3; Page 190; 444pp; German.  
XX  
XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for generation  
CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
CC sequence tags) from a particular tissue type before comparison of  
CC expression patterns. This allows a significantly longer fragment of the  
CC gene to be revealed, so should reduce the number of failures associated  
CC with the fact that ESTs from different libraries may represent different  
CC parts of the same unknown gene, distorting the estimated frequency of  
CC occurrence in a particular tissue. Z41981-Z42121 represent EST fragments  
CC derived from a human endometrium tumour cDNA library which encode the  
CC protein sequences represented in Y59941-Y60328.  
XX  
SQ Sequence 692 BP; 146 A; 189 C; 197 G; 160 T; 0 other;

Query Match 5.8%; Score 30; DB 20; Length 692;  
Best Local Similarity 57.4%; Pred. No. 4.5;  
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 1 ttttaaatgctgtcttttagaagccactgcctcagcttctgagctcagataccagaagg 60  
DB 645 TTAATAAGTAGCTGACATGAGAAAAGCCCTTGTAAAAATCCGGGGCAAAAGTGACCAGAGGA 586  
QY 61 agctggtacacagcatgataaaagacaatggga 94  
DB 585 AGACTTCTTGAGCTCTCTTCAAGGCCAAGGGA 552

Search completed: June 21, 2001, 18:06:24  
Job time: 13792 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:20:53 ; Search time 150.46 Seconds  
(without alignments)  
634.116 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_10094\_10608  
Perfect score: 515  
Sequence: 1 tttaaatgtgtctttag.....agggttcctattggtacc 515

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	29.6	5.7	152331	4	US-09-128-155-16
C 2	29.2	5.7	3688	6	Sequence 16, Appl
C 3	28.8	5.6	2088	4	Patent No. 5248670
C 4	28.4	5.5	2017	1	Sequence 64, Appl
C 5	28.4	5.5	2017	1	Sequence 27, Appl
C 6	28.4	5.5	2017	1	Sequence 27, Appl
C 7	28.4	5.5	2017	1	Sequence 27, Appl
C 8	28.2	5.5	410	4	Sequence 27, Appl
C 9	28.2	5.5	751	4	Sequence 88, Appl
C 10	28.2	5.5	3773	4	Sequence 7, Appl
C 11	28	5.4	1143	2	Sequence 1, Appl
C 12	28	5.4	2369	2	Sequence 19, Appl
C 13	28	5.4	2369	3	Sequence 4, Appl
C 14	28	5.4	12588	2	Sequence 4, Appl
C 15	27.8	5.4	376	4	Sequence 1, Appl
C 16	27.4	5.3	349	1	Sequence 3, Appl
C 17	27.4	5.3	2236	3	Sequence 1, Appl
C 18	27.4	5.3	2236	4	Sequence 11, Appl
C 19	27.4	5.3	4680	1	Sequence 11, Appl
C 20	27.4	5.3	4680	1	Sequence 1, Appl
C 21	27.4	5.3	4680	2	Sequence 1, Appl
C 22	27.4	5.3	4680	5	Sequence 1, Appl
C 23	27.4	5.3	4910	2	Sequence 1, Appl
C 24	27.4	5.3	4910	2	Sequence 2, Appl
C 25	27.2	5.3	710	1	Sequence 22, Appl
C 26	27.2	5.3	710	1	Sequence 22, Appl
C 27	27.2	5.3	710	1	Sequence 22, Appl

C 28	27.2	5.3	710	1	US-08-483-554B-22	Sequence 22, Appl
C 29	27.2	5.3	710	1	US-08-488-011B-22	Sequence 22, Appl
C 30	27.2	5.3	710	4	US-08-850-727-22	Sequence 22, Appl
C 31	27.2	5.3	710	5	PCT-US95-10202-22	Sequence 22, Appl
C 32	27.2	5.3	710	5	PCT-US95-10203-22	Sequence 22, Appl
C 33	27.2	5.3	710	5	PCT-US95-10220-22	Sequence 22, Appl
C 34	27.2	5.3	2610	2	US-08-989-386-2	Sequence 1, Appl
C 35	27.2	5.3	5656	1	US-08-425-061-1	Sequence 1, Appl
C 36	27.2	5.3	5656	1	US-08-825-886-1	Sequence 1, Appl
C 37	27.2	5.3	5689	1	US-08-425-061-3	Sequence 3, Appl
C 38	27.2	5.3	5689	1	US-08-825-886-3	Sequence 3, Appl
C 39	27.2	5.3	5707	1	US-08-425-061-11	Sequence 11, Appl
C 40	27.2	5.3	5707	1	US-08-825-886-11	Sequence 11, Appl
C 41	27.2	5.3	5709	1	US-08-425-061-2	Sequence 2, Appl
C 42	27.2	5.3	5709	1	US-08-425-061-7	Sequence 7, Appl
C 43	27.2	5.3	5709	1	US-08-425-061-8	Sequence 8, Appl
C 44	27.2	5.3	5709	1	US-08-425-061-9	Sequence 9, Appl
C 45	27.2	5.3	5709	1	US-08-825-886-2	Sequence 2, Appl

## ALIGNMENTS

### RESULT 1

US-09-128-155-16/c  
; Sequence 16, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; CURRENT FILING DATE: 1998-08-03  
; EARLIER APPLICATION NUMBER: US 60/091,650  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: US 60/054,646  
; EARLIER FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 152331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(152331)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-16

Query Match	5.7%	Score 29.6;	DB 4;	Length 152331;
Best Local Similarity	51.5%	Pred. No. 41;		
Matches	68;	Conservative	0;	Mismatches 64;
				Indels 0;
				Gaps 0;
QY	158	agggaatttcattcaatcagcaatttagtcagatctgcatctctgtcttacaagaaa	217	
Db	25835	ATGCTGAAGTCATTTCTTGTGTCATGTGAAGGAGAAATGCTCTTACTATTTTAAACAGA	25776	
QY	218	tgtcaatgggctgagatcatcagatggaggttcacatcggtttcaatgtccctatcctt	277	
Db	25775	AAGGAGAGGGTTTATGGAGATTACAGACAGGATTTCTTGATAGCAATGATTGTCATCAAT	25716	
QY	278	ttgaagacctt	289	
Db	25715	TTAGAAGATTTT	25704	

### RESULT 2

5248670-4/c  
; Patent No. 5248670  
; APPLICANT: DRAPER, KENNETH G.; ECKER, DAVID J.; MTRABELLI,  
; CHRISTOPHER K.; CROOKE, STANLEY T.



```

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,327
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/434,255
FILING DATE: 03-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3764.400-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2017 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-327-27

Query Match 5.5%; Score 28.4; DB 1; Length 2017;
Best Local Similarity 54.9%; Pred. No. 9.5;
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 355 gggttgaccatctgccattctctgttatgacagagcttgtgaaacttgaactggga 414
Db 423 CTCTTGTTCAGACAGAACCTTGTGCCTTTTGGATACACCATCTTTGAATTGACGATGA 364

QY 415 ctg99gcgaagtcaatccccacccttatcacatgaattgctga 456
Db 363 CTTACCGGGATTGGAAGCGCCTTATCCAAAGTTTGTGTA 322

RESULT 7
US-08-459-871-27/c
; Sequence 27, Application US/08459871
; Patent No. 5650326
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan P.
; APPLICANT: Outtrup, Helle
; APPLICANT: Dambmann, Claus
; APPLICANT: Aaslyng, Dorrit
; TITLE OF INVENTION: ALKALINE PROTEASE
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5650326o No. 5650326disk of No. 5650326th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,871
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/434,255  
FILING DATE: 03-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Agis Dr., Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3764,400-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2017 base pairs  
TYPE: nucleic acid  
STRADEDNESS: single  
TOPOLOGY: linear  
US-08-459-871-27

Query Match 5.5%; Score 28.4; DB 1; Length 2017;  
Best Local Similarity 54.9%; Pred. No. 9.5;  
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 355 gtttggaccatcgccattctcttattgacagagcttgaaactttaactggga 414  
Db 423 GCTTGTTCAGACGAGAACCTTGTGCGCTTTTGGATACACCATCTTTGAATTGACGATGA 364

QY 415 ctggggcaagtcacatccacaccttatacaatgaattgctga 456  
Db 363 CTTACCGGATTCGAAAGCGCTTTATCCAAAGTTTGTGA 322

RESULT 8  
US-09-276-531-88  
Sequence 88, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murry, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 88:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 base pairs  
TYPE: nucleic acid  
STRADEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT08  
CLONE: 1651564  
US-09-276-531-88

Query Match 5.5%; Score 28.2; DB 4; Length 410;  
Best Local Similarity 53.1%; Pred. No. 4.5;  
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 402 actttaactgggactggggcaaaagtcacatccacaccttatacaatgaattgctgaagagg 461  
Db 71 AATCTGTGCTGCCAAGATAAAGTCAACTTCATATTATATAGTTATATATTCGCGAGA 130  
QY 462 ccttttaaaactggagtgctgcatgtttatggaagggttcttctattggatc 514  
Db 131 TCTGAGAAAGATGGACGTTACACTGATCTTCTATGTTCTTTTATTAGTTTC 183

RESULT 9  
US-09-130-242-7  
Sequence 7, Application US/09130242B  
Patent No. 6194558  
GENERAL INFORMATION:  
APPLICANT: Gianturco, S.H.  
APPLICANT: Bradley, W.A.  
TITLE OF INVENTION: DNA Encoding Human Monocyte-Macrophage Aoplipoprotein  
FILE REFERENCE: D5880  
CURRENT APPLICATION NUMBER: US/09/130,242B  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: MS WORD, Macintosh OS 8.5  
SEQ ID NO 7  
LENGTH: 751  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-130-242-7

Query Match 5.5%; Score 28.2; DB 4; Length 751;  
Best Local Similarity 54.3%; Pred. No. 6.4;  
Matches 57; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 34 agcttctgcagctcagataccacaaaggaagctctgtacacagcatgataaagacaatggg 93  
Db 285 agggcctggagatgacagaagacatgaagtggggagctcagctgtagaacagacctgggg 344  
QY 94 acgggggtcacagtggctccctcccttccaggggtatggagacga 138  
Db 345 ctgggggagatggcagctccctgggtcccaagcagagagggcagga 389

RESULT 10  
US-09-130-242-1  
Sequence 1, Application US/09130242B  
Patent No. 6194558  
GENERAL INFORMATION:  
APPLICANT: Gianturco, S.H.  
APPLICANT: Bradley, W.A.  
TITLE OF INVENTION: DNA Encoding Human Monocyte-Macrophage Aoplipoprotein  
FILE REFERENCE: D5880  
CURRENT APPLICATION NUMBER: US/09/130,242B  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: MS WORD, Macintosh OS 8.5

APPLICANT: Bandman, Olga

APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: TWO NEW WD-40 PROTEINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/204,764  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/883,534  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0332 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2369 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: NEUTG01  
CLONE: 1221143  
US-09-204-764-4

Query Match 5.4%; Score 28; DB 3; Length 2369;  
Best Local Similarity 58.3%; Pred. No. 14;  
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 363 accatctgcccattctctgttatgacagagctgtgtaacttgaactgggactgggca 422  
Db 789 ACCTTCTCCCACTCTTCCCGTCATAGATGTATATCTGCCGTCAGCACTGCTGTGGCA 730  
QY 423 aagtcatacccacatttatacaat 446  
Db 729 AATCTGTTCCTCAGGAGAGAT 706

RESULT 14  
US-08-387-942C-1/c  
Sequence 1, Application US/08387942C  
Patent No. 5939289  
GENERAL INFORMATION:  
APPLICANT: ERTESVAG, HELGA  
APPLICANT: VALLA, SVEIN  
APPLICANT: SKJAK-BRAEK, GUDMUND  
APPLICANT: LARSEN, BJORN  
TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
STREET: P.O. BOX 747  
CITY: FALLS CHURCH  
STATE: VA

COUNTRY: USA  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/387,942C  
FILING DATE: 09-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR, GERALD M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1809-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12588 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Azotobacter vinelandii  
STRAIN: E  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 290..1951  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2227..6438  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 6702..9695  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9973..12588  
US-08-387-942C-1

Query Match 5.4%; Score 28; DB 2; Length 12588;  
Best Local Similarity 55.0%; Pred. No. 36;  
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 284 gacctggaattggcaacgcaggaacacaggaactccaccctgggtgcgtgaattgcaga 343  
Db 5694 GACCTGGCTGTCTGTCGAAGCCCTGCAGCAGGACTTCCGGATAGTGGTCCCGTTCTGCGAATT 5635  
QY 344 gctgttgtgtgtgtgtgtgtgacctatgtcccatctcttcctg 383  
Db 5634 GTCGTGGATCTGGTTGTCGAGGATCTGCACGCTCCTCCGTG 5595

RESULT 15  
US-08-985-950-3/c  
Sequence 3, Application US/08985950  
Patent No. 6140076  
GENERAL INFORMATION:  
APPLICANT: Adema, Gosse Jan  
TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..374
; US-08-985-950-3

```

```

Query Match      5.4%; Score 27.8; DB 4; Length 376;
Best Local Similarity 54.4%; Pred. No. 5.9;
Matches 56; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 285 accctgaagtggcaacgaggaagaaacaggaactccaccctggcgcgtaattgcagag 344
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AACTTCAGGGGAAATAGAGAGAGAGGGGATGTCGATGGAGCCACCCCTGGACTCCAGAG 182
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 345 ctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CAGCGTTCAGGTGGTTGACCCCAAGCCATTTTCTGTGTGT 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 21, 2001, 18:21:24  
Job time: 17477 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 15:33:59 ; Search time 4579.66 Seconds  
(without alignments)  
1063.009 Million cell updates/sec

Title: US-09-445-201-1-copy\_10094\_10608

Perfect score: 515

Sequence: 1 tttaaatgtgtgtcttag.....agggttcctattggtacc 515

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

em\_esthum10:\*

44: em\_esthum11:\*

45: em\_esthum12:\*

46: em\_esthum13:\*

47: em\_esthum14:\*

48: em\_esthum15:\*

49: em\_esthum16:\*

50: em\_esthum17:\*

51: em\_esthum18:\*

52: em\_esthum19:\*

53: em\_esthum20:\*

54: em\_esthum21:\*

55: em\_esthum22:\*

56: em\_esthum23:\*

57: em\_esthum24:\*

58: em\_esthum25:\*

59: em\_esthum26:\*

60: em\_esthum27:\*

61: em\_esthum28:\*

62: em\_estin1:\*

63: em\_estin2:\*

64: em\_estin3:\*

65: em\_estin4:\*

66: em\_estin5:\*

67: em\_estom1:\*

68: em\_estom2:\*

69: em\_estov1:\*

70: em\_estov2:\*

71: em\_estpl1:\*

72: em\_estpl2:\*

73: em\_estpl3:\*

74: em\_estpl4:\*

75: em\_estpl5:\*

76: em\_estpl6:\*

77: em\_estpl7:\*

78: em\_estpl8:\*

79: em\_estpl9:\*

80: em\_estpl10:\*

81: em\_estro1:\*

82: em\_estro2:\*

83: em\_estro3:\*

84: em\_estro4:\*

85: em\_estro5:\*

86: em\_estro6:\*

87: em\_estro7:\*

88: em\_estro8:\*

89: em\_estro9:\*

90: em\_estro10:\*

91: em\_estro11:\*

92: em\_estro12:\*

93: em\_estro13:\*

94: em\_estro14:\*

95: em\_estro15:\*

96: em\_estro16:\*

97: em\_estro17:\*

98: em\_estro18:\*

99: em\_estro19:\*

100: em\_estro20:\*

101: gb\_est25:\*

102: gb\_est26:\*

103: gb\_est27:\*

104: gb\_est28:\*

105: gb\_est29:\*

106: gb\_est30:\*

107: gb\_est31:\*

108: gb\_est32:\*

109: gb\_est41:\*

110: gb\_est42:\*

111: gb\_est43:\*

112: gb\_est44:\*

113: gb\_est45:\*

114: gb\_est46:\*

115: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE
JOURNAL
COMMENT

```

1. 586
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0110F03"
/clone_lib="Mouse 10kb p
/sex="Male"

```

BASE COUNT  
ORIGIN

Query Match 31.5%; Score 162; DB 242; Length 586;  
Best Local Similarity 99.4%; Pred. No. 2.6e-39;  
Matches 162; Conservative 0; Mismatches 1; Indels 0

1	t t t a a t g t g t g t c t t t t a a g c c a c t g c t c a g t c t c t c a g c t c a g a t a c c a a g g a	60
2y		
424	T T T A A T G T G T G T T T T A A G C C A C T G C C T C A G C T T C G C A G T C A G A T A C C A A G G A	483
db		
61	a g t c t g t c a c a g c a t g a t a a a a g a c a a t g g a c g g g g t c a c a g t g g c t c c c t c c c t t	120
2y		
484	A G T C T G T T A C A C A G C A T G A T A A A A G A C A A T G G G A C G G N G T C A C A G T G G C T C C C G T C C C T T	543
db		
121	t c a g g g t a t a g a c a c a g c t g t a g a g a g a t g t c t c c a g g a g	163
2y		
544	T C A G G G T A T G G A C A C A G C T G T A G A G A G A T G T C C A G G A G	586
db		

RESULT	2
AA067448/c	
LOCUS	
DEFINITION	AA067448 409 bp mRNA EST
	26293 Lambda-PRL2 Arabidopsis thaliana cDNA clone 91F1077, mrna
ACCESSION	sequence.
VERSION	AA067448
KEYWORDS	AA067448.1 GI:1565781
	EST.

Result No.	Score	Query		DB	ID	Description	
		Match	Length				
1	162	31.5	586	242	AZ364007	AZ364007	1M0110F03
c	37.2	7.2	409	1	AA067448	AA067448	2M2933 Lam
3	37	7.2	481	234	AQ866634	AQ866634	nbe60028L
c	34.8	6.8	627	120	AQ866634	AQ866634	W775452
c	34.8	6.8	658	175	BG307433	BG307433	fml15a04.Y
c	34.8	6.8	670	153	BG456937	BG456937	NC099G02P
c	34.2	6.6	411	136	BE050560	BE050560	dc25d01.x
c	33.8	6.6	436	232	AQ694161	AQ694161	HS_2136.B
c	33.8	6.6	648	153	BG457226	BG457226	NF01A055
c	33.8	6.6	654	151	BF640846	BF640846	NF058D111
c	33.8	6.6	669	151	BF642736	BF642736	NF070F061
c	33.8	6.6	661	153	BG457415	BG457415	NF106A08P
c	33.8	6.6	666	153	BG457203	BG457203	NF095F08P
c	33.8	6.6	691	119	AW689637	AW689637	NF022E07S
c	33.6	6.5	645	151	BF637629	BF637629	NF040D03P
c	33.6	6.5	670	247	AZ556206	AZ556206	1M0531B10
c	33.4	6.5	660	120	AW776413	AW776413	EST335478
c	33.4	6.5	782	151	BF617491	BF617491	HVSMEC001
c	33.2	6.4	322	166	BE324147	BE324147	NF015A12P
c	33.2	6.4	427	166	BE317241	BE317241	NF056C03L
c	33.2	6.4	463	166	BE323738	BE323738	NF007E08P
c	33.2	6.4	553	118	AW595530	AW595530	fk31lh09.Y
c	33.2	6.4	592	120	AW776405	AW776405	EST335470
c	33.2	6.4	596	119	AW683376	AW683376	NF011D05L
c	33.2	6.4	601	151	BF638064	BF638064	NF041C05P
c	33.2	6.4	603	163	BE123966	BE123966	EST394091
c	33.2	6.4	605	143	BF005312	BF005312	EST343810
c	33.2	6.4	620	123	AW981299	AW981299	EST392452
c	33.2	6.4	623	143	BF006418	BF006418	EST3434916
c	33.2	6.4	624	165	BE249416	BE249416	NF014H12L
c	33.2	6.4	624	251	AW795869	AW795869	2M0190E11
c	33.2	6.4	625	151	BF637320	BF637320	NF074A01L
c	33.2	6.4	629	120	AW776060	AW776060	EST335125
c	33.2	6.4	631	150	BF520706	BF520706	EST3458179
c	33.2	6.4	634	165	BE249346	BE249346	NF0014D01L
c	33.2	6.4	640	120	AW776945	AW776945	EST3366010
c	33.2	6.4	644	153	BG449793	BG449793	NF053A03I
c	33.2	6.4	646	151	BF641416	BF641416	NF0631H061
c	33.2	6.4	647	120	AW776971	AW776971	EST3360306
c	33.2	6.4	648	166	BE318918	BE318918	NF004H04L
c	33.2	6.4	649	120	AW775661	AW775661	EST334726
c	33.2	6.4	652	153	BG449849	BG449849	NF053G111
c	33.2	6.4	653	143	BF006087	BF006087	EST343657
c	33.2	6.4	657	153	BG455793	BG455793	NF070C04P
c	33.2	6.4	658	119	AW693996	AW693996	NF071C09S

RESULT	1
Z3364007	
TITLE	DNA
OCUS	GSS
DEFINITION	Mouse 10kb plasmid UGCGIM library Mus musculus genomic clone UGCGIM110F03 F, DNA sequence.
ACCESSION	AZ364007
VERSION	AZ364007.1
KEYWORDS	GI:10477707
SYNOPSIS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
AUTHORS	1 (bases 1 to 586) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islami,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly, M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

SOURCE ORGANISM	REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE	COMMENT
thale cress. Arabidopsis thaliana	Newman,T., debruin,J.F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrgoe,J.J., Raikhe,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones	Plant Physiol. 106, 1241-1255 (1994)	95148729
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.				Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, Mi Tel: 517-353-0854 Fax: 517-353-9168 Email: 22313tcn@bm.cl.msu.edu Seq primer: T7 dye primer.

FEATURES	Source	Location/Qualifiers
1..409		
/organism="Arabidopsis thaliana"		
/strain="var columbica"		
/db_xref="taxon:3702"		
/clone="91F107"		
/clone_lib="Lambda-PRL2"		
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."		
104 a	86 c	94 g
		109 t
		16 others

Query Match	7.2%; Score 37.2; DB 1; Length 409;
Best Local Similarity	64.3%; Pred. No. 1.3;
Matches	54; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 176	agcaattagtcagatctgtgcacacctatgctttacaagaaatgctcagtgggcctgagat 235 
Db 360	ACCAAGTCTCGACGCTCTCTGGATCTTCAGAGCAACCAAAAATGCCGTGTGCAGTTN 301 
QY 236	catcagatgagaggttcacgcggtt 259 
Db 300	CATNAGAAGGAGGTTTCATAGGTTT 277 
RESULT 3	
AQ866634	
LOCUS	AQ866634 481 bp DNA GSS 03-NOV-1999
DEFINITION	nbe0028L19f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic clone nbe0028L19f, DNA sequence.
ACCESSION	AQ866634
VERSION	AQ866634.1 GI:6217091
KEYWORDS	GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
AUTHORS	1 (bases 1 to 481)
TITLE	Wing, R.A. and Dean, R.A.
JOURNAL	A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998)

Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: TAATACGACTCATATAGGG  
 Class: BAC ends  
 High quality sequence start: 36  
 High quality sequence stop: 412.

FEATURES

source	Location/Qualifiers
1. .481	/organism="Oryza sativa" /strain="Japonica" /cultivar="Nipponbare" /db_xref="taxon:4530" /clone="nbbe0028L19" /clone_lib="CUGI Rice BAC Library (EcoRI)" /tissue_type="Leaf" /lab_host="E. coli DH10B" /note="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;

Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate.

Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu)).

136 a 97 c 90 q 154 t  
BAC/EST  
BASE COUNT  
4 others

BASE COUNT	136 a	97 c	90 g	154 t	4 others
ORIGIN					
Query Match	7.2%; Score 37; DB 234; Length 481;				
Best Local Similarity	56.0%; Pred. No. 1.6;				
Matches	70;	Conservative	0;	Mismatches	55; Indels 0; Gaps 0;
QY	197	catcctatgctttacaagaataatgcagttggcgctgagatcatcagatggaggttcacgg	256		
Db	268	CATCCAAATTCCTTTAGAGAGACATCCAGAGGTTCTGAAGAGCCAGATCAATCTTCAGTAA	327		
QY	257	gtttcaatgtcccgatcccttttgaagacctggaacttggaacgcagcaaacacaggaa	316		
Db	328	CTATCTCTGATGCTGATATTAAATGCATCACCTTCAAAGTTTGGTGCCAGTAAAGACTCA	387		
QY	317	ctcca	321		
Db	388	CTCCA	392		
RESULT	4				
LOCUS	AW775452/c				
DEFINITION	EST334517 DSIL Medicago truncatula cDNA clone pDSIL-1H24, mRNA sequence.				
ACCESSION	AW775452				
VERSION	AW775452.1				
KEYWORDS	GI:7765265				
SOURCE	EST.				
ORGANISM	barrel medic.				
	Medicago truncatula				







```

QY 341 agagctgtgtgtgtgtgtgaccatctgccatttcttctgttatgacagagctgtg 400
    || || || || || || || || || || || || || || || || || || || ||
Db 415 AGCTCAGTGAGTGTGTAATACCCATCTTCCAAAGCCTCTTCTGTCNATTTGTCATC 356
    || || || || || || || || || || || || || || || || || || || ||
QY 401 aactttaactgggactggggcaagtcacatccacacctttatatacaatgaattgctgaag 460
    || || || || || || || || || || || || || || || || || || || ||
Db 355 TCATCTCCCAATTCAGTCTGTCGCCCTGCTGGAAAGTAATGTGATCATTTGGAGAGAA 296
    || || || || || || || || || || || || || || || || || || || ||
QY 461 gccttttaaaactggagtgatgctgtgttatggaaggcctttcc 506
    || || || || || || || || || || || || || || || || || || || ||
Db 295 GCATTCGGCTTTTGGAAATTTTCAGTGTGTTTTCATGTTGGTTTTCCT 250
    || || || || || || || || || || || || || || || || || || || ||

RESULT 9
BG457226/c 648 bp mRNA EST 19-MAR-2001
LOCUS
DEFINITION
  NF101A05PL1F1035 Phosphate starved leaf Medicago truncatula cDNA
  clone NF101A05PL 5', mRNA sequence.
ACCESSION
  BG457226
VERSION
  BG457226.1 GI:13380551
KEYWORDS
  EST.
SOURCE
  barrel medic.
ORGANISM
  Medicago truncatula
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE
  1 (bases 1 to 648)
AUTHORS
  Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores
  H.R., Inman, J.T., Weller, J.W., May, G.D., and Harrison, M.J.
  Expressed Sequence Tags from the Samuel Roberts Noble Foundation
  Medicago truncatula phosphate-starved leaf library
  Unpublished (2000)
JOURNAL
  Contact: Harrison MJ
COMMENT
  Plant Biology Division
  The Samuel Roberts Noble Foundation
  2510 Sam Noble Parkway, Ardmore, OK 73402, USA
  Tel: 580 221 7325
  Fax: 580 221 7380
  Email: mjharrison@noble.org
  Insert Length: 648 Std Error: 0.00
  Plate: 101 row: A column: 05
  Seq primer: TCACAGGAAACAGCTATGAC.
  Location/Qualifiers
    1..648
    /organism="Medicago truncatula"
    /db_xref="taxon:3880"
    /clone_lib="Phosphate starved leaf"
    /tissue_type="leaf"
    /dev_stage="trifoliolate"
    /note="Vector: Lambda Zap; At the trifoliolate stage, M.
    truncatula plants were transplanted to phosphate-free sand
    and grown for a further 30 days. During this 30 day
    period, the plants were fertilized twice weekly with 1/2
    Hoaglands solution containing only 20uM potassium
    phosphate. RNA was prepared from above ground tissues."
    BASE COUNT 191 a 162 c 133 g 161 t 1 others
    ORIGIN

Query Match 6.6%; Score 33.8; DB 153; Length 648;
Best Local Similarity 57.8%; Pred. No. 17;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 398 gtgaactttaactgggactggggcaagtcacatccacacctttatatacaatgaattgctgaa 457
    || || || || || || || || || || || || || || || || || || || ||
Db 613 GGGAGNTGAATGTGACGCGCGGTAGTCAATGCCATCTTTCTCTCAAAATTTGACAGA 554
    || || || || || || || || || || || || || || || || || || || ||
QY 458 gagggccttttaaaactggagtgatgctgttatggaagg 499
    || || || || || || || || || || || || || || || || || || || ||
Db 553 GCGCTGTGATGAACCTCAAGGGCTTCAATCTCTGTCGAG 512
    || || || || || || || || || || || || || || || || || || || ||

```

```

RESULT 10
BF640846/c 654 bp mRNA EST 19-DEC-2000
LOCUS
DEFINITION
  NF058D11N1F1093 Insect herbivory Medicago truncatula cDNA clone
  NF058D11N 5', mRNA sequence.
ACCESSION
  BF640846
VERSION
  BF640846.1 GI:11905004
KEYWORDS
  EST.
SOURCE
  barrel medic.
ORGANISM
  Medicago truncatula
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE
  1 (bases 1 to 654)
AUTHORS
  Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores
  H.R., Inman, J.T., Weller, J.W. and May, G.D.
  Expressed Sequence Tags from the Samuel Roberts Noble Foundation
  Medicago truncatula insect herbivory library
  Unpublished (2000)
JOURNAL
  Contact: Korth K
COMMENT
  Dept. of Plant Pathology
  University of Arkansas
  217 Plant Science Building, Fayetteville, AR 72701, USA
  Tel: 501 575 5191
  Fax: 501 575 7601
  Email: kkorth@comp.uark.edu
  Insert Length: 654 Std Error: 0.00
  Plate: 058 row: D column: 11
  Seq primer: TCACACGGAACAGCATGAC.
  Location/Qualifiers
    1..654
    /organism="Medicago truncatula"
    /db_xref="taxon:3880"
    /clone_lib="NF058D11N"
    /tissue_type="local and systemic leaves"
    /dev_stage="mature"
    /note="Vector: Lambda Zap; Library was produced from fully
    expanded M. truncatula leaves of plants fed upon by
    Spodoptera exigua (beet armyworm) for 24 hours. Systemic
    (undamaged leaves from injured plants) and wounded leaves
    were harvested and pooled."
    BASE COUNT 194 a 163 c 131 g 165 t 1 others
    ORIGIN

Query Match 6.6%; Score 33.8; DB 151; Length 654;
Best Local Similarity 57.8%; Pred. No. 17;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 398 gtgaactttaactgggactggggcaagtcacatccacacctttatatacaatgaattgctgaa 457
    || || || || || || || || || || || || || || || || || || || ||
Db 603 GGGAGNTGAATGTGACGCGCGGTAGTCAATGCCATCTTTCTCTCAAAATTTGACAGA 544
    || || || || || || || || || || || || || || || || || || || ||
QY 458 gagggccttttaaaactggagtgatgctgttatggaagg 499
    || || || || || || || || || || || || || || || || || || || ||
Db 543 GCGCTGTGATGAACCTCAAGGGCTTCAATCTCTGTCGAG 502
    || || || || || || || || || || || || || || || || || || || ||

RESULT 11
BF642756/c 659 bp mRNA EST 19-DEC-2000
LOCUS
DEFINITION
  NF070F06IN1F1058 Insect herbivory Medicago truncatula cDNA clone
  NF070F06IN 5', mRNA sequence.
ACCESSION
  BF642756
VERSION
  BF642756.1 GI:11906914
KEYWORDS
  EST.
SOURCE
  barrel medic.
ORGANISM
  Medicago truncatula
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

```

REFERENCE 1 (bases 1 to 659)  
 AUTHORS Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores  
 ,H.R., Inman,J.T., Weller,J.W. and May,G.D.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 Medicago truncatula insect herbivory library  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Korth K  
 Dept. of Plant Pathology  
 University of Arkansas  
 217 Plant Science Building, Fayetteville, AR 72701, USA  
 Tel: 501 575 5191  
 Fax: 501 575 7601  
 Email: korth@comp.uark.edu  
 Insert Length: 659 Std Error: 0.00  
 Plate: 070 row: F column: 06  
 Seq primer: TCACACAGGAACAGCTATGAC.

FEATURES  
 source  
 1..659  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF070F06IN"  
 /clone\_lib="Insect herbivory"  
 /tissue\_type="local and systemic leaves"  
 /dev\_stage="mature"  
 /note="Vector: Lambda Zap; Library was produced from fully  
 expanded M. truncatula leaves of plants fed upon by  
 Spodoptera exigua (beet armyworm) for 24 hours. Systemic  
 (undamaged leaves from injured plants) and wounded leaves  
 were harvested and pooled."

BASE COUNT 198 a 160 c 133 g 166 t 2 others  
 ORIGIN

Query Match 6.6%; Score 33.8; DB 151; Length 659;  
 Best Local Similarity 57.8%; Pred. No. 17;  
 Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
 QY 398 gtgaactttaactggagtgcaattgttatggaagg 499  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 589 GGGGAAGTGAACGTGACTGCGGCTAGTCAATGCCATCTTCTCCCTCAAAATTTGACAGA 530  
 QY 458 gaggccttttaaaactggagtgcaattgttatggaagg 499  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 529 GCCGCTCGATGAACCTCAAAAGGGCTCCTCAATCTCGTCGAG 488

RESULT 12  
 BG457415/c  
 LOCUS BG457415 661 bp mRNA EST 19-MAR-2001  
 DEFINITION NF106A08PL1F1055 Phosphate starved leaf Medicago truncatula cDNA  
 clone NF106A08PL 5', mRNA sequence.  
 ACCESSION BG457415  
 VERSION BG457415.1 GI:13380824  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
 1 (bases 1 to 661)  
 AUTHORS Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores  
 ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 Medicago truncatula phosphate-starved leaf library  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Harrison MJ  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7325  
 Fax: 580 221 7380  
 Email: mjharrison@noble.org  
 Insert Length: 661 Std Error: 0.00

FEATURES  
 source

1..661  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF106A08PL"  
 /clone\_lib="Phosphate starved leaf"  
 /tissue\_type="leaf"  
 /dev\_stage="trifoliolate"  
 /note="Vector: Lambda Zap; At the trifoliolate stage, M.  
 truncatula plants were transplanted to phosphate-free sand  
 and grown for a further 30 days. During this 30 day  
 period, the plants were fertilized twice weekly with 1/2  
 Hoaglands solution containing only 20uM potassium  
 phosphate. RNA was prepared from above ground tissues."

BASE COUNT 198 a 158 c 137 g 163 t 5 others  
 ORIGIN

Query Match 6.6%; Score 33.8; DB 153; Length 661;  
 Best Local Similarity 56.2%; Pred. No. 17;  
 Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
 QY 395 ctgtgaactttaactggagtgcaattgttatggaagg 454  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 651 CCNGGNAAGTGAACGTGACTGCGGCTAGTCAATGCCATCTTCTCCCTCAAAATTTGAC 592  
 QY 455 gaagagccttttaaaactggagtgcaattgttatggaagg 499  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 591 AGAGCGCTCTGATGAACACNTCAAAAGGGCTCCTCAATCTCGTCGAG 547

RESULT 13  
 BG457203/c  
 LOCUS BG457203 666 bp mRNA EST 19-MAR-2001  
 DEFINITION NF095F08PL1F1073 Phosphate starved leaf Medicago truncatula cDNA  
 clone NF095F08PL 5', mRNA sequence.  
 ACCESSION BG457203  
 VERSION BG457203.1 GI:13380528  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
 1 (bases 1 to 666)  
 AUTHORS Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores  
 ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 Medicago truncatula phosphate-starved leaf library  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Harrison MJ  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7325  
 Fax: 580 221 7380  
 Email: mjharrison@noble.org  
 Insert Length: 666 Std Error: 0.00  
 Plate: 095 row: F column: 08  
 Seq primer: TCACACAGGAACAGCTATGAC.

FEATURES  
 source  
 1..666  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF095F08PL"  
 /clone\_lib="Phosphate starved leaf"  
 /dev\_stage="trifoliolate"  
 /tissue\_type="leaf"

/note="Vector: Lambda Zap; At the trifoliolate stage, M.  
 truncatula plants were transplanted to phosphate-free sand  
 and grown for a further 30 days. During this 30 day

period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 20uM potassium phosphate. RNA was prepared from above ground tissues."

```

BASE COUNT      196 a      166 c      137 g      165 t      2 others
ORIGIN

Query Match      6.6%; Score 33.8; DB 153; Length 666;
Best Local Similarity 57.8%; Pred. No. 18;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 398 gtgaactttaactgggactgggcaaaagtcacatccaccctttatatacaatgaattgctgaa 457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 606 GGGAAAGTGAACGTGACTGCGCGTAGTCAATGCCATCTTCTCCTCAAAATTGACAGA 547
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 458 gagcccttttaaaactggagtgcattgtttatgaagg 499
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 546 GCCGTCGTGAACCTCAAGGCTCCTTCAATCTCGTCGAG 505
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AW689627/c      691 bp      mRNA      EST      15-JUN-2000
LOCUS
DEFINITION      NF022E07ST1000 Developing stem Medicago truncatula cDNA clone
                  NF022E07ST 5', mRNA sequence.
ACCESSION      AW689627
VERSION
KEYWORDS
SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 691)
REFERENCE
AUTHORS      He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
              ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
              ,R.A.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula stem library
Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 691 Std Error: 0.00
Plate: 022 row: E column: 07
Seq primer: TCACACAGGAAACAGCTATGAC.
Location/Qualifiers
1..691
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF022E07St"
/clone_lib="Developing stem"
/tissue_type="stem"
/dev_stage="pooled developmental"
/note="Vector: Lambda zap; Contains a mixture of
intermodal stem segments"
INTERNAL ID# 196 a 166 c 137 g 175 t 2 others

BASE COUNT      207 a      168 c      139 g      175 t      2 others
ORIGIN

Query Match      6.6%; Score 33.8; DB 119; Length 691;
Best Local Similarity 57.8%; Pred. No. 18;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 398 gtgaactttaactgggactgggcaaaagtcacatccaccctttatatacaatgaattgctgaa 457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 GGGAAAGTGAACGTGACTGCGCGTAGTCAATGCCATCTTCTCCTCAAAATTGACAGA 586
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 458 gagcccttttaaaactggagtgcattgtttatgaagg 499
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 585 GCCGTCGTGAACCTCAAGGCTCCTTCAATCTCGTCGAG 544
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
BF637629/c      645 bp      mRNA      EST      19-DEC-2000
LOCUS
DEFINITION      NF040D03PL1F1028 Phosphate starved leaf Medicago truncatula cDNA
                  clone NF040D03PL 5', mRNA sequence.
ACCESSION      BF637629
VERSION
KEYWORDS
SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 645)
REFERENCE
AUTHORS      Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
              ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula phosphate-starved leaf library
Unpublished (2000)
Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Insert Length: 645 Std Error: 0.00
Plate: 040 row: D column: 03
Seq primer: TCACACAGGAAACAGCTATGAC.
Location/Qualifiers
1..645
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF040D03PL"
/clone_lib="Phosphate starved leaf"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/note="Vector: Lambda zap; At the trifoliolate stage, M.
truncatula plants were transplanted to phosphate-free sand
and grown for a further 30 days. During this 30 day
period, the plants were fertilized twice weekly with 1/2
Hoaglands solution containing only 20uM potassium
phosphate. RNA was prepared from above ground tissues."
BASE COUNT      191 a      160 c      130 g      162 t      2 others
ORIGIN

Query Match      6.5%; Score 33.6; DB 151; Length 645;
Best Local Similarity 57.1%; Pred. No. 20;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 395 cttgtgaactttaactgggactgggcaaaagtcacatccaccctttatatacaatgaattgct 454
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 CCNGGGAGCTGACTGTGACTGCGGCTAGTCAATGCCATCTTCTCCTCAAAATTGAC 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 455 gaagagcccttttaaaactggagtgcattgtttatgaagg 499
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 AGAGCCCTCTGTGATGAACCTCAAGGCTCCTTCAATCTCGTCGAG 495
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: June 21, 2001, 15:34:03
Job time: 7647 sec

```



Fri Jun 22 10:02:14 2001

us-09-445-201-1-copy\_6036\_6959.rge

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:37:34 ; Search time 8015.43 Seconds  
(without alignments)  
1783.086 Million cell updates/sec

Title: US-09-445-201-1-COPY\_6036\_6959  
Perfect score: 924  
Sequence: 1 gaagttcaaacaccgaatgt.....tgccagcgcgaggtgcagg 924

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_ba3:  
4: gb\_in1:  
5: gb\_in2:  
6: gb\_in3:  
7: gb\_om:  
8: gb\_ov:  
9: gb\_pat1:  
10: gb\_pat2:  
11: gb\_ph:  
12: gb\_pl1:  
13: gb\_pl2:  
14: gb\_pl3:  
15: gb\_pl4:  
16: em\_ba1:  
17: em\_ba2:  
18: em\_fun:  
19: em\_htgo\_hum:  
20: em\_htgo\_inv:  
21: em\_htgo\_rod:  
22: em\_htg\_hum1:  
23: em\_htg\_hum2:  
24: em\_htg\_hum3:  
25: em\_htg\_hum4:  
26: em\_htg\_hum5:  
27: em\_htg\_hum6:  
28: em\_htg\_hum7:  
29: em\_htg\_hum8:  
30: em\_htg\_inv1:  
31: em\_htg\_inv2:  
32: em\_htg\_other:  
33: em\_htg\_rod:  
34: em\_hum1:  
35: em\_hum2:  
36: em\_hum3:  
37: em\_hum4:  
38: em\_hum5:  
39: em\_hum6:  
40: em\_hum7:  
41: em\_in:  
42: em\_om:  
43: em\_or:

|||||gtgggc 855  
GAGAACTGGGC 915

44: em\_ov:  
45: em\_pat:  
46: em\_ph:  
47: em\_pl:  
48: em\_ro:  
49: em\_sts:  
50: em\_sy:  
51: em\_un:  
52: em\_vi:  
53: gb\_sts1:  
54: gb\_sts2:  
55: gb\_sts3:  
56: gb\_sy:  
57: gb\_un:  
58: gb\_vil:  
59: gb\_vil2:  
60: gb\_htg1:  
61: gb\_htg2:  
62: gb\_htg3:  
63: gb\_htg4:  
64: gb\_htg5:  
65: gb\_htg6:  
66: gb\_htg7:  
67: gb\_htg8:  
68: gb\_htg9:  
69: gb\_htg10:  
70: gb\_htg11:  
71: gb\_htg12:  
72: gb\_htg13:  
73: gb\_htg14:  
74: gb\_htg15:  
75: gb\_htg16:  
76: gb\_htg17:  
77: gb\_htg18:  
78: gb\_htg19:  
79: gb\_htg20:  
80: gb\_htg21:  
81: gb\_htg22:  
82: gb\_htg23:  
83: gb\_htg24:  
84: gb\_htg25:  
85: gb\_pr1:  
86: gb\_pr2:  
87: gb\_pr3:  
88: gb\_pr4:  
89: gb\_pr5:  
90: gb\_pr6:  
91: gb\_pr7:  
92: gb\_pr8:  
93: gb\_pr9:  
94: gb\_ro1:  
95: gb\_ro2:  
96: gb\_in4:  
97: gb\_pri0:  
98: em\_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	924	100.0	939	94	AF153057 Mus muscu
2	924	100.0	12845	9	AX002124 Sequence
3	441.2	47.7	550	94	X89777 M.musculus
4	220.6	23.9	4790	95	S53103 NYK/FLK-1-
5	169	18.3	5470	94	AR068047 Sequence
6	169	18.3	5470	94	XM0842 M.musculus
7	164.4	17.8	4016	95	U93307 Rattus norv
8	164.4	17.8	5892	95	U93306 Rattus norv

```

9      161      17.4      5391      94      MMFLK1      X59397 Mouse Flk-1
10     154.6     16.7      5406      9      AR005213 Sequence
11     154.6     16.7      5406      9      AR071705 Sequence
12     154.6     16.7      5406      10     I25171 Sequence 5
13     154.6     16.7      5406      10     I40602 Sequence 5
14     110      11.9     1267      93      HSDNAKDR X89776 H.sapiens D
15     72       7.8      214348    66     AC021220 Homo sapi
16     71.4      7.7       5830     88     AF035121 Homo sapi
17     70.6      7.6       2264     9      AR030848 Sequence
18     70.6      7.6       2264     10     I82807 Sequence 16
19     50        5.4      3009      9      AX060542 Sequence
20     47        5.1     41622     3      SCD25
21     46.4      5.0      393       7      AF083095 Felis cat
22     44        4.8     16074     68     AC023682 Drosophila
23     43.8      4.7      11958     1      AE004596 Pseudomon
24     42.4      4.6       1837     1      AF010151 Pseudomon
25     42.4      4.6      76201    85     AC005182 Homo sapi
26     42        4.5     168798    63     AC012683 Homo sapi
27     42        4.5     179579    74     AC069546 Homo sapi
28     41.6      4.5     178907    77     AC087726 Homo sapi
29     40.8      4.4     130278    86     AC005664 Homo sapi
30     40.6      4.4     191073    82     AP000752 Homo sapi
31     40.6      4.4     204842    82     AP001855 Homo sapi
32     40.2      4.4     155092    62     AC012073 Homo sapi
33     40        4.3      3897      8      CHK33A1 M11392 Chicken his
34     39.8      4.3     181715    78     AC090442 Homo sapi
35     39.4      4.3     5992      85     AB046862 Homo sapi
36     39.4      4.3     6636     91     AY004227 Homo sapi
37     39.4      4.3     7766     91     AY004226 Homo sapi
38     39.4      4.3     8069     89     AF311855 Homo sapi
39     39.4      4.3     8788     88     AF082075 Homo sapi
40     39.4      4.3     142300    66     AC020929 Homo sapi
41     39.4      4.3     163024    79     AL157939 Homo sapi
42     39.4      4.3     164652    66     AC021625 Homo sapi
43     39.4      4.3     166684    62     AC012386 Homo sapi
44     39.4      4.3     207375    81     AL392043 Homo sapi
45     39.2      4.2     87664     92     HS1048E9 Z99714 Human DNA s

```

## ALIGNMENTS

```

RESULT 1
AF153057 939 bp DNA ROD 22-MAY-2000
LOCUS Mus musculus tyrosine kinase FLK-1 gene, promoter and mRNA, partial
DEFINITION sequence.
ACCESSION AF153057
VERSION AF153057.1 GI:7963622
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 939)
AUTHORS Kappel,A., Risau,W. and Breier,G.
TITLE Prerequisite role of SCL/Flk-1, GATA and Ets transcription factor
in binding sites for the in vivo function of Flk-1 gene regulatory
elements
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 939)
AUTHORS Kappel,A., Risau,W. and Breier,G.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Molecular Cell Biology, Max-Planck,
Institut fuer Physiologische und Klinische Forschung, Parkstrasse
1, Bad Nauheim 61231, Germany
FEATURES
source Location/Qualifiers
1..939
/organism="Mus musculus"
/strain="129Sv"
/db_xref="taxon:10090"
/tissue_type="liver"
/dev_stage="fetus"

```

```

promoter 1..640
mRNA 641..>939
/product="tyrosine kinase FLK-1"
BASE COUNT 188 a 271 c 286 g 194 t
ORIGIN
Query Match 100.0%; Score 924; DB 94; Length 939;
Best Local Similarity 100.0%; Pred. No. 2e-232;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gaagtcaaacacgaaatgtcttcttaggggtaatacgttaacttcggagacatttaagtt 60
DB 16 GAAGTTCACACGGAATGCTTCTTAGGGTAATACAGGTAACTTCGGACGATTAAAGTT 75
QY 61 gccaatgacacagaaaaacagtagagcgttggcaacctgataaagcgcctactctaa 120
DB 76 GCAGATGACGAGAAAACAGTAGAGCGGTGGCAACCTGGTAAGCGCTATCTCTAA 135
QY 121 ttaaacattcagacggggcggtggtggtggtggtggtggtggtggtggtggtggtggt 180
DB 136 TTTAAACATTACAGCGGGGGGATGCGGTGGCCAAAGCACCATATAACAAACATTC 195
QY 181 agtactgacaaactcaactgcaagtgttgcccccaggtacatctaggttcagggttctt 240
DB 196 AAGTACTGACCAACTCAGTGCAGGTTTGTGCCCGGATACATCTAGGTTTCAGGGTCTT 255
QY 241 gcttcactcctcaactcggcggtggtggtggtggtggtggtggtggtggtggtggtggt 300
DB 256 GTCCTCATCTCTCCAACTCGGGCGGATTTTGTGCTCCCTTGGGACTTTCAGTGCAGG 315
QY 301 gaagagagttctgcaactcagctcctcaatgaaggcgagtggtggtggtggtggtggtggt 360
DB 316 GAAGAGAGTTCTGCAGTTCTGCAGGCTCCTAATAGGGCGCAGTGGGCTCGTGTTCGT 375
QY 361 gatgttccaggttctgctggggcgagcaagtgtctcagagccattactggtacatttt 420
DB 376 GATGCTTCCAGGTTCTGGGGGACGCAAGTCTCTAGAGCCCATTAATGGCTACATTTT 435
QY 421 acttcaccagaaacagctcgctcagatttgcctcagatgcgactgcccgcgcgcgcgcgc 480
DB 436 ACTTCCACAGAACCGAGCTGCTCCAGATTGCTCTCAGTCCGACTTGCCTCCCGCCGC 495
QY 481 acagttccgggttagtgggggagtggtggtggtggtggtggtggtggtggtggtggtggt 540
DB 496 ACAGTTCCGGGTAGTGGGGAGTGGGGAGTGGGAAACCGGAAACCGGAAACCGGAAAC 555
QY 541 agtggggcggtggcgagcagagagagagagagagagagagagagagagagagagagagag 600
DB 556 AGTGGGGGCGTGGCGGACGAGGAGTCCCGGCTCCCGGTAATGACCCCGCCCGCC 615
QY 601 attcgctagtgtagccgctctcttctgctcctgagtcctcagagaccccaagagag 660
DB 616 ATTCCGTAGTGTACCGCGGCTCTCTTCTGCCCTGAGTCTCTCAGGACCCCAAGAGAG 675
QY 661 taagctgtttctcttagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
DB 676 TAAGCTGTGTTCTTAGATCGCGGACGCGCTACCGCGGAGGACTGAAAGCCGAGACTG 735
QY 721 tctccgcagcgggataacctgctgacccgattccgcgagacacccgctgcagcgcgcgcgc 780
DB 736 TGTCCTGACCGCGGGATTAACCTGGCTGACCCGATTCGCGGACACCGCTGCGAGCCGCG 795
QY 781 tggagcagggcgccggtgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
DB 796 TGGAGCAGGGCGCGGTGCGCGGCTCTCCCGGCTCTCCCGGCTCTTCCGCTGCGGGGCGC 840
QY 841 gctctgtgactctcttggggcgagagagagagagagagagagagagagagagagagagag 900
DB 856 GCTCTGTGACTTCTTGGCGGCGCAGGAGGAGGAGTCTCTGTGCGGCTCTCTGTGCGG 924
QY 901 tctgtgcccagcgcgaggtgcag 924

```

Db 916 TCTGTGCCCCAGCGGAGGTGCAGG 939

RESULT 2  
AX002124  
LOCUS AX002124 12845 bp DNA PAT 10-MAR-2000  
DEFINITION Sequence 1 from Patent WO9855638.  
ACCESSION AX002124  
VERSION AX002124.1 GI:7241839  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 12845)  
AUTHORS Breier,G. and Risau,W.  
TITLE REGULATORY SEQUENCES CAPABLE OF CONFERRING EXPRESSION OF A  
HETEROLOGOUS DNA SEQUENCE IN ENDOTHELIAL CELLS IN VIVO AND USES  
THEREOF  
JOURNAL Patent: WO 9855638-A 1 10-DEC-1998;  
MAX PLANCK GESELLSCHAFT (DE); BREIER GEORG (DE)  
FEATURES  
Location/Qualifiers  
1..12845  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 3321 a 2893 c 2863 g 3723 t 45 others  
ORIGIN

Query Match 100.0%; Score 924; DB 9; Length 12845;  
Best Local Similarity 100.0%; Pred. No. 1.3e-232; Indels 0; Gaps 0;  
Matches 924; Conservative 0; Mismatches 0;

QY 1 gaagttcaacacgaatgtctttagggcctaatacaggttaacttcgacgatttaagt 60  
DB 6036 GAAGTTCACACCGAATGCTCTTAGGGCTAATCAGGTAACCTCGACGAGTTAAGTT 6095  
QY 61 gccagatgacagagaaaaacagtagagcgcttggaacctggataagcgcctatctctaa 120  
DB 6096 GCCAGATGACGAGAAAACAGTAGAGCGCTTGGCACTGGATAAGCGCTATCTTCTAA 6155  
QY 121 ttaaacattcagagggcggggggagtcggtggcgaagaccataaaacaaactcc 180  
DB 6156 TTAACAACTTACAGCGGGCGGGGATGCGGTGGCCAAAGCACCATAAAACAACTCC 6215  
QY 181 aagttactgcaactcaactcgaagttgtgccccagtagtacatctaggttcagggttttt 240  
DB 6216 AAGTACTGACCAACTCACTGCAAGTTGTGCCCGAGTACATCTAGTTTCAGGGTCTT 6275  
QY 241 gtcttcagtcctcccaactgcgagcgagttttgtcccttggaactttcagtcgagcggc 300  
DB 6276 GTCTTCATGCTCCCACTCGGGCGGATTTTGTGGTCCCTTGGGACTTTCAGTGCAGCGC 6335  
QY 301 gaagagagttctgacttgaggctcctcaatgaagggcgagtcggcctcggtttctggt 360  
DB 6336 GAAGAGAGTTCTGCACTTGCAGGCTCCTAATCAGGCGCAGTGGCCCTGCTGTTCTGTT 6395  
QY 361 gatgttccccaggttgcctg999ggcgagagtgctctcagagcccatctactggtacatttt 420  
DB 6396 GATGTTTCCAGGTTGCTG999ggcgagagtgctctcagagcccatctactggtacatttt 6455  
QY 421 acttcaccagaaacccagctgcctccagatttgcctcagatgcgactgcccgcggc 480  
DB 6456 ACTTCACCAGAAACCGAGCTCGCTCCAGATTGTCTCAGATGGGACTTTCAGTGCAGCGC 6515  
QY 481 acagttccggggtagtg999gggagtg999gggaaacgggaaaccccaaaccttggtatcc 540  
DB 6516 ACAGTTCGGGGTAGTGGGGGAGTGGGGTGGGAAACCGGAAACCCAAACCTTGGTATCC 6575  
QY 541 aatggggggcggtg999ggcgagcagggagtcctcccccctcccggttaatgaccccgcccc 600  
DB 6576 AGTGGGGGCGGTGGCGGACGAGGAGTCCCCACCCCTCCCGGTAAATGACCCCGCCCC 6635  
QY 601 attcgtcagtggtgtagcggcgctctcttctgtccctcagtcagtcagcccaagagag 660

Db 6636 ATTGCGTAGTGTAGCGCGGCTCTCTTCTGCGCTAGTCTCAGACCCCAAGAGAG 6695  
QY 661 taagctgtgttcttagtagtcg999ggagccgctaccg999caggactgaaagcccgactg 720  
DB 6696 TAAAGCTGTGTTTCTTATAGTCCGCGGACCGCTACCCGGCAGGACTGAAAGCCAGACTG 6755  
QY 721 tgtcccgagcgggataaactgtgactgacccgattccg999gagacacccgctgagcgcggc 780  
DB 6756 TGTCCCGCAGCGGGATAAAGTGTGCTGACCGGATTCGCGGACACCGCTGACGCGCGGC 6815  
QY 781 tggagccagggcgccggtgcccgcgctctcccgcgctcttgcgctg999gggagcctacc 840  
DB 6816 TGGAGCCAGGCGCGCGTGCCTCCCGCGCTCTCCCGGCTTTCGCGTGGGGGCGCATACC 6875  
QY 841 gcctctgtgactcttgcg999cagggcagggcaggaagagtcgtgctgagaaactgggc 900  
DB 6876 GCCTCTGTGACTTCTTTTGGCGGCGCAGGAGGAGGAGTCTGTGCTGAGAACTGGC 6935  
QY 901 tctgtgcccagcgcgaggtgcagg 924  
DB 6936 TCTGTGCCCCAGCGGAGGTGCAGG 6959

RESULT 3  
MMDNAKDR  
LOCUS MMDNAKDR 550 bp DNA ROD 10-DEC-1995  
DEFINITION M.musculus DNA for 5' UTR of receptor tyrosine.  
ACCESSION X89777  
VERSION X89777.1 GI:1019394  
KEYWORDS KDR/flk-1 gene.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 550)  
AUTHORS Patterson,C., Perrella,M.A., Hsieh,C.M., Yoshizumi,M., Lee,M.E. and  
Haber,E.  
TITLE Cloning and functional analysis of the promoter for KDR/flk-1, a  
receptor for vascular endothelial growth factor  
J. Biol. Chem. 270 (39), 23111-23118 (1995)  
MEDLINE 96032749  
REFERENCE 2 (bases 1 to 550)  
AUTHORS Patterson,C.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUL-1995) C. Patterson, Harvard School of Public  
Health, Cardiovascular Biology Lab, 677 Huntington Avenue, Building  
2 Room 113a, Boston MA 02115, USA  
FEATURES  
Location/Qualifiers  
1..550  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
promoter 1..295  
/gene="KDR/flk-1"  
gene 1..550  
/gene="KDR/flk-1"  
CDS 504..>550  
/gene="KDR/flk-1"  
/codon\_start=1  
/product="vascular endothelial growth factor (VEGF)  
receptor"  
protein\_id="CAA61917.1"  
/db\_xref="GI:1129146"  
/db\_xref="SWISS-PROT:P35918"  
/translation="MESKALLAVALWFCV"  
BASE COUNT 91 a 173 c 185 g 101 t  
ORIGIN

Query Match 47.7%; Score 441.2; DB 94; Length 550;  
Best Local Similarity 98.4%; Pred. No. 2.1e-105;  
Matches 498; Conservative 0; Mismatches 3; Indels 5; Gaps 5;









## RESULT 9

MMFLK1 5391 bp mRNA ROD 06-NOV-1991  
 Mouse Flk-1 mRNA for a tyrosine kinase receptor.  
 X59397  
 VERSION 1 GI:50976  
 KEYWORDS Flk-1 gene; tyrosine kinase receptor.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 5391)  
 Jordan,C.T.  
 Direct Submission  
 Submitted (09-MAY-1991) C.T. Jordan, Princeton University, Dept.  
 Molecular Biology, Washington Rd, Princeton, New Jersey 08544, USA  
 2 (bases 1 to 5391)  
 Matthews,W., Jordan,C.T., Gavin,M., Jenkins,N.A., Copeland,N.G. and  
 Lemischka,I.R.  
 A receptor tyrosine kinase cDNA isolated from a population of  
 enriched primitive hematopoietic cells and exhibiting close genetic  
 linkage to c-kit  
 Proc. Natl. Acad. Sci. U.S.A. 88 (20), 9026-9030 (1991)  
 92020984 Location/Qualifiers  
 1..5391  
 /organism="Mus musculus"  
 /strain="C3H/He"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 1..5391  
 /gene="Flk-1"  
 /evidence=experimental  
 1..5391  
 /gene="Flk-1"  
 208..4311  
 /gene="Flk-1"  
 /codon\_start=1  
 /product="tyrosine kinase receptor"  
 /protein\_id="CAA42040.1"  
 /db\_xref="GI:50977"  
 /db\_xref="MGD:MG1:95555"  
 /db\_xref="SWISS-PROT:P35918"  
 /translation="MESKALLAVLWFCVETRAASVGLPGDFLHPKLSKQDKILITIL  
 ANTIQTCRQDLDLWPNQRDSEERVLTECGGDSIPCKTLTIPRVVGNDTGA  
 YKSYRDYDIASVYVYRDYKSPFIASVSDHGIYITENKNTWIPCRGSIENLN  
 VSLCARYPEKRFVDPGNRISWDSEIGFTLPSTMIYAGMVFCEAKINDETVOSIMIV  
 VVGVYRIDVILSPHEIELSAGEKLVNLTARTLNVLDFETHSPSKSHKKIIV  
 RDYKFPFGTAKMFLSTLITISVKSDDGETCVASSGRMKNRRTFVRVHTKPIAF  
 GSGMSLVEATVGSVRIPVLYLSYPADPDIKWYRNPISNYTMIVGDELTIMETVE  
 RDAGNYTILNPTISMEKSHMVSILVNVPPQIGEKALISPMDSYQYGTMTLTCTVY  
 ANPPLHIIQWYMOLEACSYRPGQTSYACKERHVEDFQGNKLEVTKNQYALIEGK  
 NKTSTVLIOANYSALYKCAINKAGRGVERISPHVIRGPETVQPAAPQTEQESVS  
 LLCTADRNTFENLWYKIGSOATSVHMGESLTPVCKNLDAWLKNGTFESNTDILI  
 VAFONASLQGDYVCSAQDKTKRKHGLVQLIILERMAMPINLENQNTTIGETI  
 EVTCPASGNPTHTWFKDNETLVDSGIVLRDGNRLTIRVRKDEGLYTCQACNV  
 LGCARAEFTLIEGAQERLNEVLIVGTAVTAMFFLLVILVTVTRANGSELKTG  
 YLSIVMDPELPLDCERLPLDASKFEPRDRKLKPLGRGAFQVIEADAFDIDK  
 TATCTKTVAKMLKCATSEHRALMSKILIHIGHILNVNLLGACFKPGPLAVIV  
 EFSKFNLSLTVLGRNREFVYKSKGAFROGKDYVGLSVDLKRFLDSITSSQSSAS  
 SGFVEEKSLSDEVEEASEELYKDFLTLHLICYSFQVAKMEFLASRCLHRDLAAR  
 NILLSEKNVVICDFGLARDYKDPYVYKGDARPLKMAPETIFDRVYITQSDWS  
 FGLVLEWIFSLGASYPGVKIDEEFCRLKEGTRMAPDYTPTEMQTMIDCWHDEN  
 QRPSFSELVHLNLLQAAQDKDYIVLPMSETLSMEEDSGLSLPTSPVSCMBEE  
 VCDPKFHYDNTAGISHLQNSKRSPVSVKTFEDIPLEPEVKVIPPDSOTDSGMVL  
 ASEELKLTLEDNRNKLSPFGMMPSKRSVASEGNSOTSGYQSGVHSDDTDTTVYSSD  
 EAGLLKMWDAVHADSGTLTLQITSLNGSGVPAPPPTPGNHERGA"  
 208..264  
 /gene="Flk-1"  
 265..4308  
 /gene="Flk-1"  
 /product="tyrosine kinase receptor"  
 1397 a 1299 c 1422 g 1273 t

## FEATURES

source Location/Qualifiers

## mRNA

## gene

## CDS

## ORIGIN

Query Match 17.4%; Score 161; DB 94; Length 5391;  
 Best Local Similarity 98.1%; Pred. No. 7.2e-32;  
 Matches 205; Conservative 0; Mismatches 0; Indels 4; Gaps 4;  
 QY 718 ctgtgtcccgagccgggataacctggtgacccgattccgcgacacaccgtgcagccgc 777  
 Db 1 CTGTGTCCCGCAGCC -GGATAACCTGGCTGACCCGATTCGCCGACACCGCTGCAGCCGC 59  
 QY 778 ggtcggagccagggcgccggtgcgcccgctctcccggcttctgcgtcgggggcgcat 837  
 Db 60 GGCTGGAGCCAGGCGCGGCTGCCCGCTCTCCCGGTCTTCCGCTGCGGGGCG -CAT 118  
 QY 838 accgcctctgtgactcttcttcggggccaggagcaggaagagtcgtgctctgag-aact 896  
 Db 119 ACCGCTCTGTGACTTCTTTTCGGGGCCAGGACGAGAGAGTCTGTGCTGAGAACT 178  
 QY 897 gggctctgtgccca-gcgcgaggtgcagg 924  
 Db 179 GGCTCTGTGCCACAGGCGCGAGGTGCAGG 207

## RESULT 10

AR005213  
 LOCUS AR005213 5406 bp DNA PAT 04-DEC-1998  
 DEFINITION Sequence 5 from patent US 5747651.  
 ACCESSION AR005213  
 VERSION AR005213.1 GI:3966092  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 5406)  
 AUTHORS Lemischka,I.R.  
 TITLE Antibodies against tyrosine kinase receptor flk-1  
 JOURNAL Patent: US 5747651-A 5 05-MAY-1998;  
 FEATURES Location/Qualifiers  
 source 1..5406  
 /organism="unknown"  
 BASE COUNT 1412 a 1298 c 1423 g 1273 t  
 ORIGIN

Query Match 16.7%; Score 154.6; DB 9; Length 5406;  
 Best Local Similarity 96.2%; Pred. No. 3.5e-30;  
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;  
 QY 718 ctgtgtcccgagccgggataacctggtgacccgattccgcgacacaccgtgcagccgc 777  
 Db 1 CTGTGTCCCGCAGCC -GGATAACCTGGCTGACCCGATTCGCCGACACCGCTGCAGCCGC 59  
 QY 778 ggtcggagccagggcgccggtgcgcccgctctcccggcttctgcgtcgggggcgcat 837  
 Db 60 GGCTGGAGCCAGGCGCGGCTGCCCGCTCTCCCGGTCTTCCGCTGCGGGGCGCAT 118  
 QY 838 accgcctctgtgactcttcttcggggccaggagcaggaagagtcgtgctctgag-aact 896  
 Db 119 ACCGCTCTGTGACTTCTTTTCGGGGCCAGGACGAGAGTCTGTGCTGAGAACT 178  
 QY 897 gggctctgtgccca-gcgcgaggtgcagg 924  
 Db 179 GGCTCTGTGCCACAGGCGCGAGGTGCAGG 207

## RESULT 11

AR071705  
 LOCUS AR071705 5406 bp DNA PAT 18-FEB-2000  
 DEFINITION Sequence 5 from patent US 5912133.  
 ACCESSION AR071705  
 VERSION AR071705.1 GI:7222593  
 KEYWORDS

## sig\_peptide

## mat\_peptide

## BASE COUNT

1397 a 1299 c 1422 g 1273 t

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5406)  
AUTHORS Lemischka, I.R.  
TITLE Method for isolating stem cells expressing flk-1 receptors  
JOURNAL Patent: US 5912133-A 5 15-JUN-1999;  
FEATURES Location/Qualifiers  
1..5406  
/organism="unknown"  
BASE COUNT 1412 a 1298 c 1423 g 1273 t  
ORIGIN

Query Match 16.7%; Score 154.6; DB 9; Length 5406;  
Best Local Similarity 96.2%; Pred. No. 3.5e-30;  
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactggctgacccgattcccgagacacccgtcagccgc 777  
Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCCGATTCGCGGACACACCCGTGCGCGCC 59

QY 778 ggctggagccagggcgccggtgcccgcgctctccccggtcttcccggtctgctgctgagcgcat 837  
Db 60 GGCTGGAGCCAGGCGCGCGGTG-CCCGCGCTCTCCCGGTCTTGGCGTGGCGGGGCGCGAT 118

QY 838 accgctctgtgactcttcttggcgccagggagcaggaagagtgctgtgctgag-aact 896  
Db 119 ACCGCTCTGTGACTCTTTTGGCGGCCAGGACGGAGAGGAGTCTGTGCTGAGAAACT 178

QY 897 ggctctgtgcccc-gcgcgaggtgcagg 924  
Db 179 GGGCTCTGTGCCCCAGGCGCGAGGTGCAGG 207

RESULT 12  
125171  
LOCUS 125171 5406 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 5 from patent US 5348065.  
ACCESSION 125171  
VERSION 125171.1 GI:1605041  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5406)  
AUTHORS Lemischka, I.R.  
TITLE Tyrosine kinase receptor human flk-2-specific antibodies  
JOURNAL Patent: US 5548065-A 5 20-AUG-1996;  
FEATURES Location/Qualifiers  
1..5406  
/organism="unknown"  
BASE COUNT 1412 a 1298 c 1423 g 1273 t  
ORIGIN

Query Match 16.7%; Score 154.6; DB 10; Length 5406;  
Best Local Similarity 96.2%; Pred. No. 3.5e-30;  
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactggctgacccgattcccgagacacccgtcagccgc 777  
Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCCGATTCGCGGACACACCCGTGCGCGCC 59

QY 778 ggctggagccagggcgccggtgcccgcgctctccccggtcttcccggtctgctgctgagcgcat 837  
Db 60 GGCTGGAGCCAGGCGCGCGGTG-CCCGCGCTCTCCCGGTCTTGGCGTGGCGGGGCGCGAT 118

QY 838 accgctctgtgactcttcttggcgccagggagcaggaagagtgctgtgctgag-aact 896  
Db 119 ACCGCTCTGTGACTCTTTTGGCGGCCAGGACGGAGAGGAGTCTGTGCTGAGAAACT 178

QY 897 ggctctgtgcccc-gcgcgaggtgcagg 924

Db 179 GGGCTCTGTGCCCCAGGCGCGAGGTGCAGG 207

RESULT 13  
140602  
LOCUS 140602 5406 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 5 from patent US 5621090.  
ACCESSION 140602  
VERSION 140602.1 GI:2082894  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5406)  
AUTHORS Lemischka, I.R.  
TITLE Nucleic acids encoding soluble human FLK-2 extracellular domain  
JOURNAL Patent: US 5621090-A 5 15-APR-1997;  
FEATURES Location/Qualifiers  
1..5406  
/organism="unknown"  
BASE COUNT 1412 a 1298 c 1423 g 1273 t  
ORIGIN

Query Match 16.7%; Score 154.6; DB 10; Length 5406;  
Best Local Similarity 96.2%; Pred. No. 3.5e-30;  
Matches 20; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaactggctgacccgattcccgagacacccgtcagccgc 777  
Db 1 CTGTGTCCCGCAGCC-GGATAAAGCTGGCTGACCCGATTCGCGGACACACCCGTGCGCGCC 59

QY 778 ggctggagccagggcgccggtgcccgcgctctccccggtcttcccggtctgctgctgagcgcat 837  
Db 60 GGCTGGAGCCAGGCGCGCGGTG-CCCGCGCTCTCCCGGTCTTGGCGTGGCGGGGCGCGAT 118

QY 838 accgctctgtgactcttcttggcgccagggagcaggaagagtgctgtgctgag-aact 896  
Db 119 ACCGCTCTGTGACTCTTTTGGCGGCCAGGACGGAGAGGAGTCTGTGCTGAGAAACT 178

QY 897 ggctctgtgcccc-gcgcgaggtgcagg 924  
Db 179 GGGCTCTGTGCCCCAGGCGCGAGGTGCAGG 207

RESULT 14  
HSDNAKDR  
LOCUS HSDNAKDR 1267 bp DNA PAT 14-JAN-1996  
DEFINITION H.sapiens DNA for 5'UTR of receptor tyrosine.  
ACCESSION X89776  
VERSION X89776.1 GI:1155008  
KEYWORDS KDR/flk-1 gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1267)  
AUTHORS Patterson,C., Perrella,M.A., Hsieh,C.M., Yoshizumi,M., Lee,M.E. and Haber,E.  
TITLE Cloning and functional analysis of the promoter for KDR/flk-1, a receptor for vascular endothelial growth factor  
JOURNAL J. Biol. Chem. 270 (39), 23111-23118 (1995)  
MEDLINE 96032749  
REFERENCE 2 (bases 1 to 1267)  
AUTHORS Patterson,C.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUL-1995) C. Patterson, Harvard School of Public Health, Cardiovascular Biology Lab, 677 Huntington Avenue, Building 2 Room 113a, Boston MA 02115, USA  
REMARK Revised by [2]  
REFERENCE 3 (bases 1 to 1267)  
AUTHORS Patterson,C.

[illegible]

\* 9845 9944: gap of unknown length  
 \* 9945 14101: contig of 4157 bp in length  
 \* 14102 14201: gap of unknown length  
 \* 14202 21279: contig of 7078 bp in length  
 \* 21280 21379: gap of unknown length  
 \* 21380 31346: contig of 9967 bp in length  
 \* 31347 31446: gap of unknown length  
 \* 31447 40526: contig of 9080 bp in length  
 \* 40527 40626: gap of unknown length  
 \* 40627 52862: contig of 12236 bp in length  
 \* 52863 52963: gap of unknown length  
 \* 52964 66090: contig of 13128 bp in length  
 \* 66091 66190: gap of unknown length  
 \* 66191 82659: contig of 16469 bp in length  
 \* 82660 82759: gap of unknown length  
 \* 82760 104797: contig of 22038 bp in length  
 \* 104798 104897: gap of unknown length  
 \* 104898 129319: contig of 24422 bp in length  
 \* 129320 129420: gap of unknown length  
 \* 129421 176511: contig of 47092 bp in length  
 \* 176512 176612: gap of unknown length  
 \* 176613 177666: contig of 1055 bp in length  
 \* 177667 177767: gap of unknown length  
 \* 177768 178997: contig of 1231 bp in length  
 \* 178998 179097: gap of unknown length  
 \* 179098 180242: contig of 1145 bp in length  
 \* 180243 180342: gap of unknown length  
 \* 180343 181442: contig of 1100 bp in length  
 \* 181443 181542: gap of unknown length  
 \* 181543 182642: contig of 1100 bp in length  
 \* 182643 182742: gap of unknown length  
 \* 182743 184165: contig of 1423 bp in length  
 \* 184166 184265: gap of unknown length  
 \* 184266 185112: contig of 1247 bp in length  
 \* 185113 185612: gap of unknown length  
 \* 185613 187121: contig of 1509 bp in length  
 \* 187122 187221: gap of unknown length  
 \* 187222 188524: contig of 1303 bp in length  
 \* 188525 189782: gap of unknown length  
 \* 189783 189882: contig of 1158 bp in length  
 \* 189883 191312: contig of 1430 bp in length  
 \* 191313 191412: gap of unknown length  
 \* 191413 192837: contig of 1425 bp in length  
 \* 192838 192938: gap of unknown length  
 \* 192939 194117: contig of 1180 bp in length  
 \* 194118 194217: gap of unknown length  
 \* 194218 195948: contig of 1731 bp in length  
 \* 195949 196049: gap of unknown length  
 \* 196050 197744: contig of 1696 bp in length  
 \* 197745 197844: gap of unknown length  
 \* 197845 199489: contig of 1645 bp in length  
 \* 199490 199590: gap of unknown length  
 \* 199591 201304: contig of 1715 bp in length  
 \* 201305 201404: gap of unknown length  
 \* 201405 203317: contig of 1912 bp in length  
 \* 203318 203416: gap of unknown length  
 \* 203417 204490: contig of 1074 bp in length  
 \* 204491 204590: gap of unknown length  
 \* 204591 206676: contig of 2086 bp in length  
 \* 206677 206776: gap of unknown length  
 \* 206777 208803: contig of 2027 bp in length  
 \* 208804 208903: gap of unknown length  
 \* 208904 210597: contig of 1694 bp in length  
 \* 210598 210697: gap of unknown length  
 \* 210698 212097: contig of 1400 bp in length  
 \* 212098 212197: gap of unknown length  
 \* 212198 214348: contig of 2151 bp in length.

FEATURES  
 source

1. .214348  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"

/clone="RP11-530117"  
 1. .1584  
 /note="assembly\_name:Contig100"  
 1685. .4194  
 /note="assembly\_name:Contig101"  
 4295. .6424  
 /note="assembly\_name:Contig102"  
 6525. .9844  
 /note="assembly\_name:Contig103"  
 9945. .14101  
 /note="assembly\_name:Contig104"  
 14202. .21279  
 /note="assembly\_name:Contig105"  
 21380. .31346  
 /note="assembly\_name:Contig106"  
 31447. .40526  
 /note="assembly\_name:Contig107"  
 clone\_end:77  
 vector\_side:left  
 40627. .52862  
 /note="assembly\_name:Contig108"  
 52963. .66090  
 /note="assembly\_name:Contig109"  
 66191. .82659  
 /note="assembly\_name:Contig110"  
 82760. .104797  
 /note="assembly\_name:Contig111"  
 104898. .129319  
 /note="assembly\_name:Contig112"  
 clone\_end:SP6  
 vector\_side:right  
 129420. .176511  
 /note="assembly\_name:Contig113"  
 176612. .177666  
 /note="assembly\_name:Contig133"  
 17767. .178997  
 /note="assembly\_name:Contig137"  
 179098. .180242  
 /note="assembly\_name:Contig156"  
 180343. .181442  
 /note="assembly\_name:Contig159"  
 181543. .182642  
 /note="assembly\_name:Contig166"  
 182743. .184165  
 /note="assembly\_name:Contig169"  
 184266. .185512  
 /note="assembly\_name:Contig175"  
 185613. .187121  
 /note="assembly\_name:Contig177"  
 187222. .188524  
 /note="assembly\_name:Contig178"  
 188625. .189782  
 /note="assembly\_name:Contig180"  
 189883. .191312  
 /note="assembly\_name:Contig185"  
 191413. .192837  
 /note="assembly\_name:Contig186"  
 192938. .194117  
 /note="assembly\_name:Contig187"  
 194218. .195948  
 /note="assembly\_name:Contig188"  
 196049. .197744  
 /note="assembly\_name:Contig189"  
 197845. .199489  
 /note="assembly\_name:Contig190"  
 199590. .201304  
 /note="assembly\_name:Contig191"  
 201405. .203316  
 /note="assembly\_name:Contig193"  
 203417. .204490

Query Match 7.8%; Score 72; DB 66; Length 214348;  
 Best Local Similarity 67.3%; Pred. No. 1.1e-08;







OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:06:24 ; Search time 346.98 Seconds  
(without alignments)  
1554.597 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_6036\_6959  
 Perfect score: 924  
 Sequence: 1 gaagttcaacaccgaaatgt.....tqcccaqcqcgaqgttcacg 924

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters:	1356552
----------------------------------------------------	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Database :
N_Geneseq_0401.*
1: /SID$2/cgcdata/geneseq/geneseqn/NA1980.DAT.*
2: /SID$2/cgcdata/geneseq/geneseqn/NA1981.DAT.*
3: /SID$2/cgcdata/geneseq/geneseqn/NA1982.DAT.*
4: /SID$2/cgcdata/geneseq/geneseqn/NA1983.DAT.*
5: /SID$2/cgcdata/geneseq/geneseqn/NA1984.DAT.*
6: /SID$2/cgcdata/geneseq/geneseqn/NA1985.DAT.*
7: /SID$2/cgcdata/geneseq/geneseqn/NA1986.DAT.*
8: /SID$2/cgcdata/geneseq/geneseqn/NA1987.DAT.*
9: /SID$2/cgcdata/geneseq/geneseqn/NA1988.DAT.*
10: /SID$2/cgcdata/geneseq/geneseqn/NA1989.DAT.*
11: /SID$2/cgcdata/geneseq/geneseqn/NA1990.DAT.*
12: /SID$2/cgcdata/geneseq/geneseqn/NA1991.DAT.*
13: /SID$2/cgcdata/geneseq/geneseqn/NA1992.DAT.*
14: /SID$2/cgcdata/geneseq/geneseqn/NA1993.DAT.*
15: /SID$2/cgcdata/geneseq/geneseqn/NA1994.DAT.*
16: /SID$2/cgcdata/geneseq/geneseqn/NA1995.DAT.*
17: /SID$2/cgcdata/geneseq/geneseqn/NA1996.DAT.*
18: /SID$2/cgcdata/geneseq/geneseqn/NA1997.DAT.*
19: /SID$2/cgcdata/geneseq/geneseqn/NA1998.DAT.*
20: /SID$2/cgcdata/geneseq/geneseqn/NA1999.DAT.*
21: /SID$2/cgcdata/geneseq/geneseqn/NA2000.DAT.*
22: /SID$2/cgcdata/geneseq/geneseqn/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			Length	DB	ID	Description
	Score	Match	%				
1	924	100.0	12845	20	V84274	Mouse endothelial	
2	374.2	40.5	500	18	T60766	Mouse KDR/flik-1 protein	
3	169	18.3	5409	15	Q64049	Sequence of murine	
4	154.6	16.7	5404	17	T38735	Murine foetal liver	
5	154.6	16.7	5404	20	X77516	Murine flik-1 cDNA	
6	154.6	16.7	5406	13	Q29957	flik-1 cDNA sequence	
7	154.6	16.7	5406	14	Q33504	Murine flik-1 cDNA	
8	154.6	16.7	5406	14	Q35251	Human flik-1 coding	
9	154.6	16.7	5406	14	Q40916	Murine flik-1 cDNA	
10	154.6	16.7	5406	16	Q81014	Flik1 receptor protein	
11	154.6	16.7	5406	16	Q79070	Mouse flik-1 cDNA	

12	154.6	16.7	5406	18	T721119
13	107.6	11.6	1267	18	T60765
14	70.6	7.6	2264	15	Q74274
15	70.6	7.6	2264	19	V09331
16	70.6	7.6	2264	20	X04433
17	59.8	6.5	493	18	T60763
18	40.2	4.4	114955	20	X53491
c 19	40	4.3	1120	18	T75034
c 20	40	4.3	3937	18	T75036
21	40	4.3	3937	21	261238
22	39	4.2	1026	15	O60691
c 23	38.8	4.2	597	20	Z33423
24	38.6	4.2	15672	12	Q10613
c 25	37.4	4.0	1053	8	N70994
c 26	37.4	4.0	1053	15	Q73489
c 27	37.2	4.0	6002	20	X60227
28	36.8	4.0	13204	20	Z32189
29	36.6	4.0	1022	20	Z19442
30	36.6	4.0	1022	20	Z19230
31	36.4	3.9	58857	21	A58471
32	36.2	3.9	1938	22	C83145
33	36.2	3.9	2118	22	C83152
34	36.2	3.9	2620	21	A14342
35	36.2	3.9	2620	21	A13380
c 36	36	3.9	1640	12	Q10990
c 37	36	3.9	1641	14	Q49931
c 38	36	3.9	1641	15	Q45224
c 39	36	3.9	2393	12	Q10907
c 40	36	3.9	2394	12	C83951
41	35.6	3.9	1316	14	Q53210
c 42	35.6	3.9	114955	20	X53491
43	35.4	3.8	8438	15	Q73500
c 44	35	3.8	7898	20	X5080
45	35	3.8	10732	21	A10594

## ALIGNMENTS

RESULT	1	
IID	V84274	
XX	V84274 standard; DNA; 12845 BP.	
XX	V84274;	
XX	AC	
XX	DT	
XX	12-APR-1999 (first entry)	
XX	Mouse endothelial growth factor receptor-2 gene Flk-1 5' region.	
XX	Flk-1; endothelial growth factor receptor-2; VEGF; transcription;	
XX	endothelium; enhancer; vascular disease; angiogenesis; cancer;	
XX	diabetic retinopathy; rheumatoid arthritis; wound healing;	
XX	vulnary; atherosclerosis; tumour; neuronal disorder; therapy;	
XX	diagnosis; mouse; ss.	
OS	Mus sp.	
XX		
Key	Location/Qualifiers	
promoter	6036..6959	
FT	/*tag= a	
FT	6661..7026	
FT	/*tag= b	
FT	/number= 1	
FT	/codon_start= 6960	
FT	7027..10632	
FT	/*tag= c	
FT	/number= 1	
FT	/note= "DNA sequences comprising nucleotides	
FT	8260-10560, 8336-10608 and/or 10094-10608	
FT	are specifically claimed in Claim 3b"	
FT	protein_bind	10137..10155
FT	/*tag= d	
FT	/note= "GATA/PEA3 consensus binding site"	

protein\_bind 10166..10179  
 /tag= e  
 /note= "GATA consensus binding site"  
 10187..10197  
 /tag= f  
 /note= "AP1 consensus binding site"  
 10252..10261  
 /tag= g  
 /note= "NFKB consensus binding site"  
 10273..10283  
 /tag= h  
 /note= "AP1 consensus binding site"  
 10301..10309  
 /tag= i  
 /note= "STAT consensus binding site"  
 10356..10372  
 /tag= j  
 /note= "ETS-1/GATA consensus binding site"  
 10395..10404  
 /tag= k  
 /note= "ETS-1 consensus binding site"  
 10453..10477  
 /tag= l  
 /note= "SCL/TAL-1 Ets-1 consensus binding site"  
 10643..10726  
 /tag= m  
 /number= 2  
 10727..12673  
 /tag= n  
 /number= 2  
 12674..12845  
 /tag= o  
 /number= 3  
 WO9855638-A1.  
 10-DEC-1998.  
 03-JUN-1998; 98WO-EP03318.  
 03-JUN-1997; 97EP-0108959.  
 (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 Breier G, Risau W, Roenicke V;  
 WPI; 1999-059915/05.

New recombinant DNA containing heterologous sequence controlled by endothelial cell specific regulator - from the Flk-1 gene, used to treat, prevent or diagnose vascular disease, tumours, also to screen for transcriptional regulators  
 Claim 3; Fig 1; 107pp; English.  
 This 12.8 kb DNA sequence spans the region from -6.65 kb relative to the transcriptional start site to +6.15 (located in the third exon) of the murine endothelial growth factor (VEGF) receptor-2 gene Flk-1. This portion of the Flk-1 gene was isolated from the liver of mouse strain 129/SvJ by screening a phage library of liver genomic DNA in vector lambda FixII. Sequences within the 5'-flanking region of the gene, in combination with sequences located within the first intron, specifically and reproducibly target expression of heterologous DNA to angioblasts during early stages of vascular development and also to the vasculature of postnatal mice. The regulatory sequences of the first intron also function as an autonomous endothelium-specific enhancer when fused to a heterologous promoter (e.g. HSV-TK promoter). This Flk-1 intron enhancer contains several potential binding sites for transcription factors of the Ets and GATA families. The invention provides a new claimed recombinant DNA (I) comprising at least one regulatory sequence from an intron of the Flk-1 gene, or its homologue, to control expression in endothelial cells, in vivo, linked to (b) a

heterologous DNA. (I) is used to produce transgenic animals, and these, or transformed cells, are used to identify agents (A), potential pharmaceuticals, that suppress, activate or enhance transcription of genes in endothelial cells. (I), vectors and (A) are used to direct or prevent (for antisense sequences) expression of genes specifically in endothelial cells, e.g. for treating angiogenesis, cancer, diabetic retinopathy, rheumatoid arthritis etc., and in wound healing, particularly of smooth muscle cells, specifically atherosclerosis), tumours and neuronal disorders. They may also be used to induce vascular disease in the transgenic animals or diagnostically, particularly for studying (mal)function, interactions and unregulated expression of endothelial products. (I) provide specific modulation of gene expression in endothelial cells, at all stages of development.  
 XX Sequence 12845 BP; 3321 A; 2893 C; 2863 G; 3723 T; 45 other;  
 SQ  
 Query Match 100.0%; Score 924; DB 20; Length 12845;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-252;  
 Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gaagttcaaacacgaatgtctcttagggcgtatcaggttaactcggacgatttaaaagt 60  
 Db 6036 gaagttcaaacacgaatgtctcttagggcgtatcaggttaactcggacgatttaaaagt 6095  
 QY 61 gccagatggacagagaaacacagtagagcgcttgcaacctgataagcgctattcttaa 120  
 Db 6096 gccagatggacagagaaacacagtagagcgcttgcaacctgataagcgctattcttaa 6155  
 QY 121 ttaaacattcagacggcgggcggtggtggccaaagcaccataaaacaaactcc 180  
 Db 6156 ttaaacattcagacggcgggcggtggtggccaaagcaccataaaacaaactcc 6215  
 QY 181 aagtactgacaaactcaactcaagttgtgtcccgagatcatatcaggttcagggttctt 240  
 Db 6216 aagtactgacaaactcaactcaagttgtgtcccgagatcatatcaggttcagggttctt 6275  
 QY 241 gtcttcacgtcccaactcggcggtggtgtgtcccttgggacttcagtcagcgccg 300  
 Db 6276 gtcttcacgtcccaactcggcggtggtgtgtcccttgggacttcagtcagcgccg 6335  
 QY 301 gaagagagttctcacttcaggtcctcctaatagagggcgagtggtggcctctgtttctggt 360  
 Db 6336 gaagagagttctcacttcaggtcctcctaatagagggcgagtggtggcctctgtttctggt 6395  
 QY 361 gatgtcccgaggtgtgtggtggcgagcaagtgtctcagagcccaattactggtacattt 420  
 Db 6396 gatgtcccgaggtgtgtggtggcgagcaagtgtctcagagcccaattactggtacattt 6455  
 QY 421 acttccaccagaaacccagctgcgtccagatttgcctcagatgcgacttgcccgccggc 480  
 Db 6456 acttccaccagaaacccagctgcgtccagatttgcctcagatgcgacttgcccgccggc 6515  
 QY 481 acagttccgggttagtggggagtggtggcggtgggaaacccgggaaacccaaactggtacc 540  
 Db 6516 acagttccgggttagtggggagtggtggcggtgggaaacccgggaaacccaaactggtacc 6575  
 QY 541 agtggggcggtggcgccgagcagggagtgctccaccctcccggttaatgaccccgccccc 600  
 Db 6576 agtggggcggtggcgccgagcagggagtgctccaccctcccggttaatgaccccgccccc 6635  
 QY 601 attcgctagtgtgtagccggtcgtctcttctgtccctgagtcctcagggaccccaagagag 660  
 Db 6636 attcgctagtgtgtagccggtcgtctcttctgtccctgagtcctcagggaccccaagagag 6695  
 QY 661 taagctgtttctcctagatcgccgagcagctaccgcgagactgaaagccagactg 720  
 Db 6696 taagctgtttctcctagatcgccgagcagctaccgcgagactgaaagccagactg 6755  
 QY 721 tgtcccgagcggggataaacctggctgacccgattccgcggacacccgctgagccgccc 7800



DE Sequence of murine genomic foetal liver kinase 1 (Flk-1) gene.

XX Foetal liver kinase; receptor tyrosine kinase; ss.

XX Mus musculus.

XX Key Location/Qualifiers

XX CDS 285..4386

XX /\*tag= a

XX W09411499-A.

XX 26-MAY-1994.

XX 15-NOV-1993; 93WO-EP03191.

XX 13-NOV-1992; 92US-0975750.

XX 26-MAR-1993; 93US-0038596.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Millauer B, Risau W, Ullrich A;

XX WPI; 1994-183501/22.

XX P-PSDB; R54046.

XX DNA encoding Flk-1, a tyrosine kinase receptor for vascular endothelial growth factor - used to express recombinant Flk-1 for screening for ligands useful for modulating vasculogenesis and angiogenesis e.g. for treating cancer

XX Disclosure; Page 50-56; 99pp; English.

XX A receptor tyrosine kinase cDNA, designated foetal liver kinase 1 (Flk-1), was cloned from mouse cell populations enriched for haematopoietic stem and progenitor cells. The nucleotide coding sequence and deduced AA sequence of the murine Flk-1 gene is depicted in Q64049/R54046 and has been described in Matthews et al., 1991, Proc. Natl. Acad. Sci. USA, 88:9026-9030. The murine Flk-1 gene was isolated by performing a PCR using two degenerate oligo primer pools that were designed on the basis of highly conserved sequences within the kinase domain of receptor tyrosine kinases (Hanks et al., 1988). Based on AA homology, this receptor is a member of the type III subclass of RTKs (Ullrich and Schlessinger) which contd. immunoglobulin-like repeats in their extracellular domains.

XX Sequence 5470 BP; 1417 A; 1324 C; 1440 G; 1289 T; 0 other;

XX Query Match 18.3%; Score 169; DB 15; Length 5470;  
XX Best Local Similarity 98.2%; Pred. No. 3.8e-38;  
XX Matches 213; Conservative 0; Mismatches 4; Gaps 4;

Qy 711 gccagactgtgtccgcagcgggataaactgctgaacccgattccgcgacaccgctg 770

Db 70 gccagactgtgtccgcagcgggataaactgctgaacccgattccgcgacaccgctg 129

Qy 771 -cagccgcggtgagcagcggcggcgtgcccgcgctctcccggtcttgcgctgcg 829

Db 130 acagccgcggtgagcagcggcggcgtgcccgcgctctcccggtcttgcgctgcg 189

Qy 830 gggcgataccgctctgtgacttttgcggggccagggacggagagagtgctgtgctt 889

Db 190 gggc-cataccgctctgtgacttttgcggggccagggacggagagagtgctgtgctt 248

Qy 890 gag-aactggactgtgcccac-gcgcgagtgagg 924

Db 249 gagaaactgggctgtgtgcccagcgcgagtgagg 285

RESULT 4

T38735

ID T38735 standard; cDNA; 5404 BP.

XX

AC

XX

DT

XX

DE

XX

KW

KW

KW

KW

XX

OS

XX

XX

PH

FT

FT

FT

FT

FT

FT

FT

XX

PN

XX

XX

PD

XX

XX

XX

XX

XX

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

T38735;

11-DEC-1996 (first entry)

Murine foetal liver kinase 1 cDNA.

Murine; foetal liver kinase 1; flk-1; protein tyrosine kinase;

monoclonal; antibody; extracellular domain; receptor assay;

haematopoietic stem cell; ligand; stimulation; proliferation;

differentiation; treatment; anaemia; bone marrow damage;

cancer chemotherapy; radiation; ds.

Mus musculus.

Key Location/Qualifiers

CDS 208..4311

/\*tag= a

sig\_peptide 208..264

/\*tag= b

mat\_peptide 265..4308

/\*tag= c

US5548065-A.

20-AUG-1996.

02-APR-1991; 91US-0679666.

19-NOV-1992; 92US-0977451.

02-APR-1991; 91US-0679666.

28-JUN-1991; 91US-0728913.

15-NOV-1991; 91US-0793065.

24-DEC-1991; 91US-0813593.

26-JUN-1992; 92US-0906397.

12-NOV-1992; 92US-0975049.

30-APR-1993; 93US-0055269.

31-OCT-1994; 94US-0252517.

(UVP- ) UNIV PRINCETON.

Lemischka IR;

WPI; 1996-392678/39.

P-PSDB; T38735.

Anti-foetal liver kinase 2 (flk-2) antibodies - useful in assays,

for isolating haematopoietic stem cells expressing receptor and for

obtaining ligands

Disclosure; Columns 51-62; 50pp; English.

The present sequence encodes murine foetal liver kinase 1 (flk-1),

a protein tyrosine kinase. Isolated antibodies, pref. monoclonal,

raised against the extracellular portion of flk-1 can be used to

assay for flk receptors on the surface of haematopoietic stem

cells, and to isolate positive cells. The antibodies can also

be used as, or to obtain ligands, which stimulate the proliferation

and/or differentiation of stem cells. The ligands can be used, e.g.

for treating anaemia, or bone marrow damage resulting from cancer

chemotherapy, or radiation.

Sequence 5404 BP; 1411 A; 1297 C; 1423 G; 1273 T; 0 other;

Query Match 16.7%; Score 154.6; DB 17; Length 5404;

Best Local Similarity 96.2%; Pred. No. 4.7e-34;

Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

Qy 718 ctgtgtccgcagcgggataaactgctgacccgattccgcgacaccgctgagccgc 777

Db 1 ctgtgtccgcagcc-ggataacctgctgacccgattccgcgacaccgctgagccgc 59



CC haematopoietic cells. Ligand binding to the TK may be prep'd. which  
 CC can stimulate proliferation and/or differentiation of primitive  
 CC haematopoietic cells in vivo. The ligands can stimulate the  
 CC proliferation of additional primitive stem cells, differentiation into  
 CC more mature progenitor cells, or both.  
 CC See also Q29954-6.  
 XX  
 SQ Sequence 5406 BP; 1412 A; 1294 C; 1427 G; 1273 T; 0 other;

Query Match 16.7%; Score 154.6; DB 13; Length 5406;  
 Best Local Similarity 96.2%; Pred. No. 4.7e-34;  
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

Qy 718 ctgtgtcccgagccgggataaactggtgacccgattcccgagacacccgtgcagccgc 777  
 |||||  
 Db 1 ctgtgtcccgagcc-ggataaactggtgacccgattcccgagacacccgtgcagccgc 59  
 |||||  
 Qy 778 ggtgtggagccagggcgccggtgccccgctctcccggtcttctgctgaggggagccat 837  
 |||||  
 Db 60 ggtgtggagccagggcgccggtg-ccccgctctcccggtcttctgctgaggggagccat 118  
 |||||  
 Qy 838 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgcttgag-aact 896  
 |||||  
 Db 119 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgcttgagaaact 178  
 |||||  
 Qy 897 gggctctgtgcccc-gcgcgaggtgcagg 924  
 |||||  
 Db 179 gggctctgtgccccagcgcgaggtgcagg 207  
 |||||

RESULT 7  
 Q53504  
 ID Q53504 standard; cDNA; 5406 BP.  
 XX  
 AC Q53504;

XX 27-JUN-1994 (first entry)  
 XX Murine flk-1 cDNA.

XX Receptor protein tyrosine kinase; pTK family; foetal liver kinase;  
 KW flk1; primitive; totipotent; haematopoietic cell; stem cell;  
 KW proliferation; stromal cell; ds.  
 XX

OS Mus musculus.

XX Key Location/Qualifiers  
 FH CDS 208..4311  
 FT /\*tag= a  
 FT /product= flk-1  
 FT 208..264  
 FT /\*tag= b  
 FT /note= "hydrophobic leader"  
 FT 265..4308  
 FT /\*tag= c

XX US5270458-A.  
 XX  
 XX 14-DEC-1993.  
 XX  
 XX 02-APR-1991; 91US-0679666.  
 XX  
 XX 02-APR-1991; 91US-0679666.  
 PR 28-JUN-1991; 91US-0728913.  
 PR 15-NOV-1991; 91US-0793065.  
 PR 24-DEC-1991; 91US-0813593.  
 PR 26-JUN-1992; 92US-0906397.  
 PR 12-NOV-1992; 92US-095049.  
 PR 19-NOV-1992; 92US-0977451.

XX (UYPR-) UNIV PRINCETON.

XX

PI Lemischka IR;  
 XX  
 DR WPI; 1993-405021/50.  
 DR P-PSDB; R44996.  
 XX

XX Isolated nucleic acid molecules of hematopoietic stem cell  
 PT receptor flk-2 - encoding mammalian receptor protein tyrosine  
 PT kinases expressed in primitive haematopoietic cells  
 XX  
 XX Disclosure; Fig 2; 60pp; English.

XX Nucleic acid sequences coding for murine flk-2 and specified  
 CC subfragments of it are claimed. The murine flk-1 coding sequence  
 CC (i.e. Q53504) is also disclosed. The flk polypeptides are  
 CC receptor protein tyrosine kinases which are expressed only in  
 CC primitive haematopoietic cells.  
 XX

SQ Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;

Query Match 16.7%; Score 154.6; DB 14; Length 5406;  
 Best Local Similarity 96.2%; Pred. No. 4.7e-34;  
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

Qy 718 ctgtgtcccgagccgggataaactggtgacccgattcccgagacacccgtgcagccgc 777  
 |||||  
 Db 1 ctgtgtcccgagcc-ggataaactggtgacccgattcccgagacacccgtgcagccgc 59  
 |||||  
 Qy 778 ggtgtggagccagggcgccggtgccccgctctcccggtcttctgctgaggggagccat 837  
 |||||  
 Db 60 ggtgtggagccagggcgccggtg-ccccgctctcccggtcttctgctgaggggagccat 118  
 |||||  
 Qy 838 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgcttgag-aact 896  
 |||||  
 Db 119 accgctctgtgactttcttgcggcgccagggagcagagagagtgctgtgcttgagaaact 178  
 |||||  
 Qy 897 gggctctgtgcccc-gcgcgaggtgcagg 924  
 |||||  
 Db 179 gggctctgtgccccagcgcgaggtgcagg 207  
 |||||

RESULT 8  
 Q535251  
 ID Q535251 standard; cDNA; 5406 BP.  
 XX  
 AC Q535251;

XX 25-JUN-1993 (first entry)  
 XX Human flk-1 coding sequence.

XX Human; receptor; protein; tyrosine kinase; pTK; primitive; mammalian;  
 KW hematopoietic cell; pTK; mature; mtk; fetal; liver kinase 2; flk-1;  
 KW liver; spleen; thymus; adult; brain; marrow; thymocyte; subset;  
 KW multipotential; T-lymphoid; lineage; stomach; kidney; lung; heart;  
 KW intestine; muscle; lymph node; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 208..4311  
 FT /\*tag= a

XX WO9300349-A.  
 XX  
 XX 07-JAN-1993.  
 XX  
 XX 26-JUN-1992; 92WO-0505401.  
 XX  
 XX 28-JUN-1991; 91US-0728913.  
 PR 15-NOV-1991; 91US-0793065.  
 PR 24-DEC-1991; 91US-0813593.  
 PR 02-APR-1992; 92WO-0502750.

XX	(UYPR-) UNIV PRINCETON.
FA	Lemischka IR;
XX	
PI	WPI; 1993-036323/04.
XX	P-PSDB; R31377.
DR	
DR	
XX	Nucleic acid encoding receptor protein tyrosine kinase - allows
PT	development of ligands to stimulate proliferation and/or
PT	differentiation of mammalian haematopoietic stem cells
XX	
PS	Claim 10; Fig 2; 78pp; English.
XX	
CC	This sequence encodes a human receptor protein tyrosine kinase which
CC	belongs to a new functional class of protein tyrosine kinases (PTKs).
CC	However, this PTK is not in the same class as flk-2. PTKs in the same
CC	class as flk-2 are expressed in primitive mammalian hematopoietic
CC	(PHC) cells but not in mature hematopoietic cells (mHC). This gene
CC	is expressed in more mature hematopoietic cells. The protein encoded
CC	by this sequence is an example of a receptor PTK and is called fetal
CC	liver kinase 1 (flk-1). flk-1 is expressed in fetal liver, spleen,
CC	thymus, brain, stomach, kidney, lung, heart and intestine, and adult
CC	brain, kidney, heart, spleen, lung, muscle, marrow and lymph nodes.
XX	
SQ	Sequence 5406 BP; 1412 A; 1299 C; 1422 G; 1273 T; 0 other;
Query Match            16.7%; Score 154.6; DB 14; Length 5406;	
Best Local Similarity 96.2%; Pred. No. 4.7e-34;	
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;	
QY	718 ctgtgtcccgagcgggataaaccttgctgaccgatccgcggacacccgtcgagccgc 777
Db	
1	ctgtgtcccgagcc-ggataacctggtgaccgatccgcggacacccgtcgagccgc 59
QY	778 ggcctgaggccaggcggcggcggtgcccgcgctctccccggtcttgcgtcgggggcgcat 837
Db	
60	ggctggagccaggcggcggcggtg-cccgcgctctccccggtcttgcgtcgggggcgcat 118
QY	838 accgcctctgactcttttcggggccaggacggagagaagagtctgtgctgag-aact 896
Db	
119	accgcctctgactcttttcggggccaggacggagagaagagtctgtgctgagaaact 178
QY	897 gggtcttgtgtccca-gcgcgaggtgcagg 924
Db	
179	gggtctgtgccccaggcggcggatgcagg 207
RESULT	9
Q40916	
ID	Q40916 standard; cDNA; 5406 BP.
XX	
AC	Q40916;
XX	
DT	19-OCT-1993 (first entry)
XX	
DE	Murine flk-1 cDNA.
XX	
KW	Murine; receptor; protein; tyrosine kinase; PTK; flk-1; flk-2; adult;
KW	family; primitive; hematopoietic cell; mature; fetal; liver; spleen;
KW	thymus; brain; stomach; kidney; lung; heart; intestine; bone marrow;
KW	muscle; lymph node; ss
XX	
OS	Mus musculus.
XX	
FH	Key Location/Qualifiers
FT	CDS 208..4311
FT	/tag= a
FT	sig_peptide 208..264
FT	/tag= b
FT	/note= "Hydrophobic leader sequence"
FT	mat_peptide 265..4308

PT		/**tag= c
XX	WO9310136-A.	
PN		
XX		
PD	27-MAY-1993.	
XX		
PX	16-NOV-1992; 92WO-US09893.	
XX		
PR	15-NOV-1991; 91US-0793065.	
XX	(UYPR-) UNIV PRINCETON.	
PA		
XX	Lemisckha IR;	
PI		
XX	WPI; 1993-182479/22.	
DR	P-PSDB; R37504.	
XX		
PT	Totipotent haematopoietic stem cell receptors, their ligands and DNA sequences - for treating anaemia(s) and bone marrow damage due to e.g. cancer chemotherapy or radiotherapy	
PT		
XX	Claim 14; Fig 2; 127pp; English.	
PS		
CC	This sequence encodes the murine receptor protein tyrosine kinase (PTK), flk-1. This PTK is not in the same family as flk-2 (see also CC Q40914-15) as it is expressed in primitive hematopoietic cells and also in mature hematopoietic cells. flk-1 is expressed in fetal liver, spleen, thymus, brain, stomach, kidney, lung, heart and CC intestine and adult brain, bone marrow, kidney, spleen, lung, muscle and lymph nodes.	
XX		
SQ	Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;	
Query Match 16.7%; Score 154.6; DB 14; Length 5406; Best Local Similarity 96.2%; Pred. NO. 4.7e-34; Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps		
QY	718 ctgtgctccgcagcggataacctggctgacccgattccggcgagcacaccgtgcagccgc 777 	
Dd	1 ctgtgctccgcagcc-ggataacctggctgacccgattccggcgagcacaccgtgcagccgc 59 	
QY	778 ggcctgagccaggcgccggctgcccgcgctctccccggtcttgcgctcgggggcgcat 837 	
Dd	60 ggctgagccaggcgccggctg-cgcgctctccccggtcttgcgctcgggggccgat 118 	
QY	838 accgcctctgtactcttttcggcgccaggacggaggaaggagtctgtaccttgag-aact 896 	
Dd	119 accgcctctgtactcttttcggcgccaggacggaggaaggagtctgtacctgagaact 178 	
QY	897 gggctctgtgccc-a-gcgcgaggtgcagg 924 	
Dd	179 gggctctgtgcccaggcgaggtgcagg 207 	
RESULT	10	
Q81014		
ID	Q81014 standard; cDNA; 5406 BP.	
XX		
AC	Q81014;	
XX		
DT	18-AUG-1995 (first entry)	
XX		
DE	Flk1 receptor protein-tyrosine-kinase cDNA.	
KW	Mouse Flk1; receptor protein-tyrosine-kinase; hematopoietic stem cell; fetal liver kinase; ds.	
KW		
OS	Mus musculus.	
XX		
Key	Location/Qualifiers	
CDS	208..4311	
FT	/*tag= a	

FT		/product= Flk1 receptor protein-tyrosine-kinase	
FT	sig_peptide	208..264	
FT		/*tag= b	
FT	mat_peptide	265..4308	
FT		/*tag= c	
XX			
XX	WO9500554-A.		
PN			
XX	05-JAN-1995.		
XX			
XX	17-JUN-1994;	94WO-US06944.	
XX			
PR	18-JUN-1993;	93US-0080244.	
PR	21-JUN-1993;	93US-0081508.	
PR	23-NOV-1993;	93US-0157490.	
XX			
XX	(UYPR-) UNIV PRINCETON.		
PI			
PI	Lemischka IR;		
XX			
DR	WPI; 1995-052014/07.		
DR	P-PSDB; R67817.		
XX			
PT	Ligand for receptor protein tyrosine kinase - useful for the		
PT	stimulation of primitive haematopoietic stem cells causing		
PT	proliferation and/or differentiation		
XX			
XX	Disclosure; Fig 2: 131pp; English.		
PS			
XX	The sequence corresponds to a cDNA encoding a mouse Flk1 (fetal		
CC	liver kinase) receptor protein-tyrosine-kinase. Flk1 is expressed		
CC	in both primitive and mature hematopoietic cells, and in other		
CC	tissues. The gene product is useful in isolation of receptor		
CC	ligands, which have applications in diagnosis of bone marrow		
CC	disorders and in stimulating proliferation and/or differentiation		
CC	of primitive hematopoietic stem cells.		
XX			
SO	Sequence 5406 BP: 1412 A; 1298 C; 1423 G; 1273 T; 0 other:		

XX	Mus sp.
OS	
XX	
XX	
FH	Key Location/Qualifiers
FT	208..4311
CDS	/*tag= a
FT	208..264
FT	/*tag= b
FT	265..4308
FT	/*tag= c
XX	
XX	US5367057-A.
PN	
XX	22-NOV-1994.
PD	
XX	02-APR-1991; 91US-0679666.
PF	
XX	
XX	02-APR-1991; 91US-0679666.
PR	
PR	28-JUN-1991; 91US-0728913.
PR	15-NOV-1991; 91US-0793065.
PR	24-DEC-1991; 91US-0813593.
PR	26-JUN-1992; 92US-0906397.
PR	12-NOV-1992; 92US-0975049.
PR	19-NOV-1992; 92US-0977451.
PR	30-APR-1993; 93US-0055269.
XX	
PA	(UYPR-) UNIV PRINCETON.
PI	
XX	Lemischka IR;
XX	
DR	WPI; 1995-005894/01.
DR	P-PSDB; R67537.
XX	
PT	Murine flk-2 receptor protein tyrosine kinase - used to stimulate
PT	proliferation and/or stimulation of primitive mammalian
PT	haematopoietic stem cells in vitro or in vivo.
XX	
XX	Disclosure; Fig. 3A-3I; 69pp; English.
XX	
CC	cDNAs encoding receptor protein tyrosine-kinases, mouse fetal liver
CC	kinase-2 (flk-2), human flk-2 and mouse flk-1 are given in Q79068-70,
CC	respectively, and the deduced amino acid sequences in R67535-37,
CC	respectively.
XX	
SQ	Sequence 5406 BP; 1412 A; 1298 C; 1423 G; 1273 T; 0 other;
	Query Match 16.7%; Score 154.6; DB 16; Length 5406;
	Best Local Similarity 96.2%; Pred. No. 4.7e-34;
	Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
QY	718 ctgtgtcccgagcgggataaaccttgctgaccgcattccgcggacaccgtgcagccgc 777       Db 1 ctgtgtcccgagcc-ggataacctggtgaccgatccgcggacaccgtgcagccgc 59 
QY	778 ggctgagaccaggcgccggtgcccccgcgtctctccccgcgtcttgcgtgcggggcgcat 837 
Db	60 ggtgtgagcagagcgccggtg-ccegcgtctctccccgcgtcttgcgtgcggggcgcat 118 
QY	838 accgcctctgtactcttttcggggccagggagcaggaagaagtctgtgctgag-aact 896 
Db	119 accgcctctgtactcttttcggggccagggagcaggaagaagtctgtgctgagaaact 178 
QY	gggtctgtgtccca-gcgcgaggtgcagg 924
Db	179 gggtctgtgtccaggcgcgaggtgcagg 207 
RESULT	12
T72119	
ID	T72119 standard; cDNA; 5406 BP.
XX	
AC	T72119;











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 18:21:24 ; Search time 150.46 Seconds  
(without alignments)  
1137.715 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_6036\_6959

Perfect score: 924

Sequence: 1 gaagtcaacacccaagtgt.....tgcccagcgaggtgcagg 924

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	169	18.3	5470	2	US-08-443-861-1
2	169	18.3	5470	4	US-08-193-829B-1
3	154.6	16.7	5406	1	US-07-813-593-3
4	154.6	16.7	5406	1	US-07-977-451-5
5	154.6	16.7	5406	1	US-07-946-507-3
6	154.6	16.7	5406	1	US-08-252-517-5
7	154.6	16.7	5406	1	US-07-906-397A-5
8	154.6	16.7	5406	1	US-08-601-891-5
9	154.6	16.7	5406	2	US-09-021-324-5
10	154.6	16.7	5406	5	PCT-US92-02750-7
11	154.6	16.7	5406	5	PCT-US92-05401-5
12	154.6	16.7	5406	5	PCT-US92-09893-5
13	70.6	7.6	2264	1	US-08-232-538-16
14	70.6	7.6	2264	2	US-08-786-164-16
15	40	4.3	1120	3	US-08-586-165-1
16	40	4.3	3937	3	US-08-586-165-8
17	40	4.3	7812	4	US-09-368-590-1
18	39	4.2	1026	1	US-07-975-526-6
19	37.4	4.0	1053	6	5352575-6
20	36	3.9	1641	1	US-08-385-229-1
21	36	3.9	1641	2	US-08-650-000-1
22	36	3.9	1641	6	5395760-1
23	36	3.9	2224	4	US-08-477-347-2
24	35.4	3.8	8438	1	US-07-945-283-1
25	34.6	3.7	2010	4	US-09-240-410-1
26	34.2	3.7	1028	4	US-08-118-200-1
27	34.2	3.7	1028	4	US-08-458-745-1

Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 29, Appli  
Sequence 29, Appli  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 17, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 14, Appli  
Sequence 10, Appli  
Sequence 8, Appli

## ALIGNMENTS

RESULT 1  
US-08-443-861-1  
; Sequence 1, Application US/08443861  
; Patent No. 5851999  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Risau, Werner  
; APPLICANT: Millauer, Birgit  
; APPLICANT: Gazit, Aviv  
; APPLICANT: Levitzki, Alex  
; TITLE OF INVENTION: Flk-1 Is A Receptor For Vascular  
; TITLE OF INVENTION: Endothelial Growth Factor  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/443,861  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/193,829  
; FILING DATE: 09-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7683-060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)790-9090  
; TELEFAX: (212)869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5470 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 286...4386

US-08-443-861-1

Query Match 18.3%; Score 169; DB 2; Length 5470;  
Best Local Similarity 98.2%; Pred. No. 3.5e-36;  
Matches 213; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 711 gccagactgtgtccgcagccgggataaacctggtgacccgattccgcggacacccgctg 770  
|||||  
DB 70 GCCCAGACTGTGTCCCGCAGCGGATAACTGCTGACCCGATTCCGCGGACACCGCTG 129  
QY 771 -cagcccgctgagccagggcgccggtgcccccgctctccccggtcttctgagctg 829  
|||||  
DB 130 ACAGCCCGCTGTGAGCAGCGCGCGGTGCCCGCTCTCCCGCTCTTGGCGCTCGG 189  
QY 830 gggcgataccgctctgtgactcttctggtggcgagcgaggaagaggtctgtgctt 889  
|||||  
DB 190 GGGC-CATACCGCTCTGTGACTTCTTTGGGGCCAGGACGAGGAGTCTGTGCTT 248  
QY 890 gag-aactgggtctgtgcccc-gcgcgaggtgcag 924  
|||||  
DB 249 GAGAACTGGGCTCTGTGCCAGCGCGAGGTGCAGG 285

## RESULT 2

US-08-193-829B-1  
; Sequence 1, Application US/08193829B  
; Patent No. 6177401  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Risaau, Werner  
; APPLICANT: Millauer, Birgit  
; APPLICANT: Gazit, Aviv  
; APPLICANT: Levitzki, Alex  
; TITLE OF INVENTION: Flk-1 Is A Receptor For Vascular  
; TITLE OF INVENTION: Endothelial Growth Factor  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/193,829B  
; FILING DATE: 09-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cotuzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7683-060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)790-9090  
; TELEFAX: (212)869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5470 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 286..4386  
US-08-193-829B-1

## RESULT 3

US-07-813-593-3  
; Sequence 3, Application US/07813593  
; Patent No. 5185438  
; GENERAL INFORMATION:  
; APPLICANT: Lemischka, Ihor R.  
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED  
; STREET: 180 VARICK STREET  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10014  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/813,593  
; FILING DATE: 19920415  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/793,065  
; FILING DATE: 15-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/728,913  
; FILING DATE: 28-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/679,666  
; FILING DATE: 02-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feit, Irving N.  
; REGISTRATION NUMBER: 28,601  
; REFERENCE/DOCKET NUMBER: LEM-3-PPP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-645-1405  
; TELEFAX: 212-645-2054  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5406 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 208..4311

```
;
;
; NAME/KEY: mat_peptide
; LOCATION: 208..4308
; US-07-813-593-3

Query Match      16.7%; Score 154.6; DB 1; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagcgagataaactgactgaaccgattccgagacaccgtgagcgcg 777
Db 1 CTGTGTCCCGAGCG-GGATAAACCTGGCTACCCGATTCCCGGACACCCCGTGCGAGCGCG 59

QY 778 ggcctgtgtgcccagcgcggtgccccgcgctctcccggtcttgcgtgcggggcgcat 837
Db 60 GGCTGGAGCGAGCGCGCGGTG-CCCGCGCTCTCCCGGTCTTCCGCTGCGGGGCGCGAT 118

QY 838 accgctctgtgacttcttgcgggcccagggagagagagtgctgtgcttgag-aact 896
Db 119 ACCGCCCTCTGTGACTTCTTTTCGGGGCCAGGACGAGGAGAGTCTGTGCTGAGAAACT 178

QY 897 ggcctgtgtgcccag-cgcgaggtgcagg 924
Db 179 GGGCTCTGTGCCAGCGCGAGGTGCAGG 207

RESULT 4
US-07-977-451-5
; Sequence 5, Application US/07977451
; Patent No. 5270458
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977,451
; FILING DATE: 19921119
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US UNASSIGNED
; FILING DATE: 12-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
```

```
;
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 208..4311
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 265..4308
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 208..264
; US-07-977-451-5

Query Match      16.7%; Score 154.6; DB 1; Length 5406;
Best Local Similarity 96.2%; Pred. No. 2.5e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagcgagataaactgactgaaccgattccgagacaccgtgagcgcg 777
Db 1 CTGTGTCCCGAGCG-GGATAAACCTGGCTACCCGATTCCCGGACACCCCGTGCGAGCGCG 59

QY 778 ggcctgtgtgcccagcgcggtgccccgcgctctcccggtcttgcgtgcggggcgcat 837
Db 60 GGCTGGAGCGAGCGCGCGGTG-CCCGCGCTCTCCCGGTCTTCCGCTGCGGGGCGCGAT 118

QY 838 accgctctgtgacttcttgcgggcccagggagagagagtgctgtgcttgag-aact 896
Db 119 ACCGCCCTCTGTGACTTCTTTTCGGGGCCAGGACGAGGAGAGTCTGTGCTGAGAAACT 178

QY 897 ggcctgtgtgcccag-cgcgaggtgcagg 924
Db 179 GGGCTCTGTGCCAGCGCGAGGTGCAGG 207

RESULT 5
US-07-946-507-3
; Sequence 3, Application US/07946507
; Patent No. 5283354
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED
; STREET: 180 VARICK STREET
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/946,507  
 FILING DATE: 19920917  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/813,593  
 FILING DATE: 24-DEC-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/793,065  
 FILING DATE: 15-NOV-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/728,913  
 FILING DATE: 28-JUN-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/679,666  
 FILING DATE: 02-APR-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Feit, Irving N.  
 REGISTRATION NUMBER: 28,601  
 REFERENCE/DOCKET NUMBER: LEM-3-PPP  
 TELEPHONE: 212-645-1405  
 TELEFAX: 212-645-2054  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5406 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 208..4311  
 NAME/KEY: mat\_peptide  
 LOCATION: 208..4308  
 US-07-946-507-3

Query Match 16.7%; Score 154.6; DB 1; Length 5406;  
 Best Local Similarity 96.2%; Pred. No. 2.5e-32;  
 Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;  
 Qy 718 ctgtgtccgcagcggagataacctggtgacccgattcccgagacacgctgcagccgc 777  
 Db 1 CTGTGTCCCGCAGCC-GGATAACCTGGCTGACCCGATTCGCCGACACCCGTCGAGCCGC 59  
 Qy 778 qgcctggagcagcggcggtgcccgcgctctccccggtcttgcgctgcccggcgccat 837  
 Db 60 GGCTGGAGCCAGCGCGCGGTG-CGCCGCTCTCCCGGCTTGGCGTGGCGGGCGCGAT 118  
 Qy 838 accgcctctgtgactcttttgcggccagggagacgagagagtgctgtgctgag-aact 896  
 Db 119 ACCGCCCTGTGACTCTTTTGCGGCCAGGAGGACGAGAGAGTCTGTGCTGTGAGAACT 178  
 Qy 897 ggcctctgtgcca-gcgcaggtgtgcag 924  
 Db 179 GGGCTCTGTGCCAGCGCGGAGGTGCAGG 207

RESULT 6  
 US-08-252-517-5  
 Sequence 5, Application US/08252517  
 Patent No. 5548065  
 GENERAL INFORMATION:  
 APPLICANT: Lemischka, Ihor R.  
 TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
 TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS  
 NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: ImClone Systems Incorporated  
 STREET: 180 Varick Street  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10014  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/252,517  
 FILING DATE: 31-OCT-1994  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/977,451  
 FILING DATE: 19-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/906,397  
 FILING DATE: 26-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US92/05401  
 FILING DATE: 26-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: TW 81102961  
 FILING DATE: 15-APR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US92/02750  
 FILING DATE: 02-APR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/813,593  
 FILING DATE: 24-DEC-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/793,065  
 FILING DATE: 15-NOV-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/728,913  
 FILING DATE: 28-JUN-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/679,666  
 FILING DATE: 02-APR-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Feit, Irving N.  
 REGISTRATION NUMBER: 28,601  
 REFERENCE/DOCKET NUMBER: LEM-3-7P  
 TELEPHONE: 212-645-1405  
 TELEFAX: 212-645-2054  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5406 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 208..4311  
 NAME/KEY: mat\_peptide  
 LOCATION: 265..4308  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 208..264  
 US-08-252-517-5

Query Match 16.7%; Score 154.6; DB 1; Length 5406;





;; FILING DATE: 24-DEC-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/793,065  
;; FILING DATE: 15-NOV-1991  
;; PRIOR APPLICATION DATA: US 07/728,913  
;; FILING DATE: 28-JUN-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/679,666  
;; FILING DATE: 02-APR-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Feit, Irving N.  
;; REGISTRATION NUMBER: 28,601  
;; REFERENCE/DOCKET NUMBER: LEM-3-7P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-645-1405  
;; TELEFAX: 212-645-2054  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5406 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE: N-terminal  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 208..4311  
;; FEATURE:  
;; NAME/KEY: mat\_peptide  
;; LOCATION: 265..4308  
;; FEATURE:  
;; NAME/KEY: sig\_peptide  
;; LOCATION: 208..264  
US-08-601-891-5

Query Match 16.7%; Score 154.6; DB 1; Length 5406;

Best Local Similarity 96.2%; Pred. No. 2.5e-32;

Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataacttgctgctgacccgattccggcgagacaccgctgcagccgc 777

Db 1 CTGTGTCCCGAGCC-GGATAACCTGGCTGACCCGATTCCGGGACACCCGCTGCAGCCGC 59

QY 778 ggcctggagcagggcgccgtgcccgcgctctcccgggtcttgcgctgcggggcgcat 837

Db 60 GCGTGGAGCAGGCGCGCGGTG-CCCGCGTCTCTCCCGGTCTTGGCTGCGGGGCGCAT 118

QY 838 accgctctgtgacttttgcggcgccagggagcaggaaggagtgctgtgctgag-aact 896

Db 119 ACCGCTCTGTGACTTCTTTTCGGGCGCCAGGACGAGGAGTCTGTGCTGAGAACT 178

QY 897 gggctctgtgccc-a-gcgcaggtgcagg 924

Db 179 GGGCTCTGTGCCAGGCGCGAGGTGCAGG 207

# RESULT 9

US-09-021-324-5

Sequence 5, Application US/09021324

Patent No. 5912133

GENERAL INFORMATION:

APPLICANT: Lemischka, Ihor R.

TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL

TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: ImClone Systems Incorporated

STREET: 180 Varick Street

CITY: New York

;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10014  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/021,324  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/977,451  
;; FILING DATE: 1992-11-19  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/906,397  
;; FILING DATE: 26-JUN-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US PCT/US92/05401  
;; FILING DATE: 26-JUN-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: TW 81102961  
;; FILING DATE: 15-APR-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US PCT/US92/02750  
;; FILING DATE: 02-APR-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/813,593  
;; FILING DATE: 24-DEC-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/793,065  
;; FILING DATE: 15-NOV-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/728,913  
;; FILING DATE: 28-JUN-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/679,666  
;; FILING DATE: 02-APR-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Feit, Irving N.  
;; REGISTRATION NUMBER: 28,601  
;; REFERENCE/DOCKET NUMBER: LEM-3-7P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-645-1405  
;; TELEFAX: 212-645-2054  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5406 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE: N-terminal  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 208..4311  
;; FEATURE:  
;; NAME/KEY: mat\_peptide  
;; LOCATION: 265..4308  
;; FEATURE:  
;; NAME/KEY: sig\_peptide  
;; LOCATION: 208..264  
US-09-021-324-5

Query Match 16.7%; Score 154.6; DB 2; Length 5406;

Best Local Similarity 96.2%; Pred. No. 2.5e-32;

Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtcccgagccgggataaacctggctgacccgattccggcgagacaccgctgcagccgc 777

Db 1 CTGTGTCCTCCGAGCC-GGATAACCTGGCTGACCGGATTCGCGGACACCCGTCGACGCCG 59  
QY 778 ggcctctgtgactcttcttgcggccgagtgcccgctctcccggtcttgcgctgctgag-aact 837  
Db 60 GGCTGGAGCAGGCGCGGTG-CCGCGCTCTCTCCCGGTCTTGGCGTGGGGGGCGCAT 118  
QY 838 accgctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 896  
Db 119 ACCGCTCTGTGACTTCTTTTCGCGGCGCAGGACGGAGAGAGTCTGTGCTTGAGAACT 178  
QY 897 ggcctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 924  
Db 179 GGCTCTGTGCTCCGAGCGCGGAGGTGAGG 207

## RESULT 10

PCT-US92-02750-7

; Sequence 7, Application PC/TUS9202750

; GENERAL INFORMATION:

; APPLICANT: LEMISCHKA, THOR R.

; TITLE OF INVENTION: Totipotent Hematopoietic Stem Cell

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED

; STREET: 180 VARICK STREET

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: US

; ZIP: 10014

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/02750

; FILING DATE: 19920402

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: FEIT, IRVING N.

; REGISTRATION NUMBER: 28,601

; REFERENCE/DOCKET NUMBER: LEM-3-PPPT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-645-1405

; TELEFAX: 212-645-2054

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5406 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 208..4311

; FEATURE:

; NAME/KEY: mat\_peptide

; LOCATION: 208..4308

PCT-US92-02750-7

Query Match 16.7%; Score 154.6; DB 5; Length 5406;  
Best Local Similarity 96.2%; Pred. No. 2.5e-32;  
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtccgcagcgagggataacccgctgacccgagtgcccgagtgcccgagtgctgag-aact 777  
Db 1 CTGTGTCCTCCGAGCC-GGATAACCTGGCTGACCGGATTCGCGGACACCCGTCGACGCCG 59

QY 778 ggcctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 837  
Db 1 CTGTGTCCTCCGAGCC-GGATAACCTGGCTGACCGGATTCGCGGACACCCGTCGACGCCG 59

Db 60 GGCTGGAGCAGGCGCGGTG-CCGCGCTCTCTCCCGGTCTTGGCGTGGGGGGCGCAT 118  
QY 838 accgctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 896  
Db 119 ACCGCTCTGTGACTTCTTTTCGCGGCGCAGGACGGAGAGAGTCTGTGCTTGAGAACT 178  
QY 897 ggcctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 924  
Db 179 GGCTCTGTGCTCCGAGCGCGGAGGTGAGG 207

## RESULT 11

PCT-US92-05401-5

; Sequence 5, Application PC/TUS9205401

; GENERAL INFORMATION:

; APPLICANT: Lemischka, Ihor R.

; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL

; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: IMCLONE SYSTEMS INCORPORATED

; STREET: 180 VARICK STREET

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10014

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/05401

; FILING DATE: 19920626

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Feit, Irving N.

; REGISTRATION NUMBER: 28,601

; REFERENCE/DOCKET NUMBER: LEM-3-PPPT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-645-1405

; TELEFAX: 212-645-2054

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5406 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 208..4311

; FEATURE:

; NAME/KEY: mat\_peptide

; LOCATION: 208..4308

PCT-US92-05401-5

Query Match 16.7%; Score 154.6; DB 5; Length 5406;  
Best Local Similarity 96.2%; Pred. No. 2.5e-32;  
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 718 ctgtgtccgcagcgagggataacccgctgacccgagtgcccgagtgcccgagtgctgag-aact 777  
Db 1 CTGTGTCCTCCGAGCC-GGATAACCTGGCTGACCGGATTCGCGGACACCCGTCGACGCCG 59

QY 778 ggcctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 837  
Db 60 GGCTGGAGCAGGCGCGGTG-CCGCGCTCTCTCCCGGTCTTGGCGTGGGGGGCGCAT 118

QY 838 accgctctgtgactcttcttgcggccgagtgcccgagtgcccgagtgctgctgag-aact 896  
Db 119 ACCGCTCTGTGACTTCTTTTCGCGGCGCAGGACGGAGAGAGTCTGTGCTTGAGAACT 178

QY 897 gggtctgtgccc-gcgaggtgcagg 924  
|||||  
Db 179 GGCTCTGTGCCAGGCGGAGGTGCAGG 207

## RESULT 12

PCT-US92-09893-5  
; Sequence 5, Application PC/TUS9209893  
; GENERAL INFORMATION:  
; APPLICANT: Lemischka, Ihor R.  
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ImClone Systems Incorporated  
; STREET: 180 Varick Street  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10014  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/09893  
; FILING DATE: 19921116  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feit, Irving N.  
; REGISTRATION NUMBER: 28,601  
; REFERENCE/DOCKET NUMBER: LEM-3-7PT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-645-1405  
; TELEFAX: 212-645-2054  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5406 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 208..4311  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 265..4308  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 208..264  
PCT-US92-09893-5

Query Match 16.7%; Score 154.6; DB 5; Length 5406;  
Best Local Similarity 96.2%; Pred. No. 2.5e-32;  
Matches 201; Conservative 0; Mismatches 4; Indels 4; Gaps 4;  
QY 718 ctgtgtcccgagccgggataaacctggctgacccgattccggagacaccgtgcagcgc 777  
|||||  
Db 1 CTGTGTCCCGAGCC -GGATAACCTGCTGACCCGATTCGCGGACACCCGTGCAGCGC 59  
QY 778 ggcgtgagccagggcgccggtgccccgcgtctccccggttcttgcgtgcggggcgcat 837  
|||||  
Db 60 GGCTGGAGCCAGGCGCGGTG -CCCCGCTCTCCCCGCTCTTCGCGTGGCGGGCGCAT 118  
QY 838 accgctctgtgactcttgcggccagggacgaggaagagctctgtcctaag-aact 896  
|||||

Db 119 ACCGCTCTGTGACTTCTTTGCGGGCCAGGACGGAGAGAGAGTCTGTGCTCGAGAACT 178  
QY 897 gggtctgtgccc-gcgaggtgcagg 924  
|||||  
Db 179 GGCTCTGTGCCAGGCGGAGGTGCAGG 207

## RESULT 13

US-08-232-538-16  
; Sequence 16, Application US/08232538  
; Patent No. 5712380  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Kenneth A.  
; APPLICANT: Kendall, Richard L.  
; TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL CELL  
; TITLE OF INVENTION: GROWTH FACTOR  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: P.O. Box 2000 126 E Lincoln Avenue  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07065-0907  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,538  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wallen, John W.III  
; REGISTRATION NUMBER: 35,403  
; REFERENCE/DOCKET NUMBER: 18888IA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 594-3905  
; TELEFAX: (908) 594-4720  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2264 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-232-538-16

Query Match 7.6%; Score 70.6; DB 1; Length 2264;  
Best Local Similarity 63.9%; Pred. No. 6e-10;  
Matches 140; Conservative 0; Mismatches 74; Indels 5; Gaps 2;

QY 711 gccagactgttcccgagccgggataaacctggctgacccgattccggagacaccgtg 770  
|||||  
Db 51 GCAGAAAGTCCGTCTGGCAGCTGGATATCTCTCTACCGGACCGGACGCGCCCTG 110  
QY 771 cagccgctgtgagccagggcgccggtgccccgcgtctccccggtcttgcgtgcgg 830  
|||||  
Db 111 CAGCCGCGTTCGGCGCCGCGGCTCCCTAGCCCTGTGGCTCACTGTCTGCGTGGCGG 170  
QY 831 g----gcgcataccgcctctgtgactcttcttgcggggccaggagcgagagagtg 886  
|||||  
Db 171 GTGCGCGGAGTTCCACCTCGCGGCTCTTCTCTAGACAGGCGCTGGGAGAAAGACGG 230  
QY 887 cctgagaact-gggctgtgtcccgagcgaggtgcagg 924  
|||||  
Db 231 CTCCCGAGTTCGCGCATTTCCCGCGCTCGAGGTGCAGG 269

RESULT 14  
US-08-786-164-16

```

: Sequence 16, Application US/08786164
: Patent No. 5861484
: GENERAL INFORMATION:
: APPLICANT: THOMAS, KENNETH A.
: APPLICANT: KENDALL, RICHARD L.
: TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL
: TITLE OF INVENTION: CELL GROWTH FACTOR
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merck & Co., Inc.
: STREET: P.O. Box 2000, 126 E. Lincoln Ave.
: CITY: Rahway
: STATE: NJ
: COUNTRY: USA
: ZIP: 07065-0900
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Microsoft Word 6.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/786,164
: FILING DATE: 21-JAN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Mark Hand, J
: REGISTRATION NUMBER: 36,545
: REFERENCE/DOCKET NUMBER: 18888DA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 908-594-3905
: TELEFAX: 908-594-4720
: TELEX:
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2264 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-786-164-16

```

Query Match 7.6%; Score 70.6; DB 2; Length 2264;  
Best Local Similarity 63.9%; Pred. No. 6e-10;  
Matches 140; Conservative 0; Mismatches 74; Indels 5

Qy	711	gccagactgtctccgcagccgcgggataaacctgctgacccgattccgcggacacccgtg	770
Db	51	GCAGAAAGTCGCTGGCAGCCCTGGGATATCCTCTCTACCGGCACCCGCAGACGCCCTG	110
Qy	771	cagccgcgctgaagccagagccgcggtgccccgcgcctctcccgcgtcttgcgctcgagg	830
Db	111	CAGCCGCGTGGCGCCGGGCTCCCTAGCCCTGTGCGTCTCAACTGCTCGCTCGCGG	170
Qy	831	g----gcgcataccgcctctgtgactctcttggcgccaggagcgcgagagagtgctgtg	886
Db	171	GTGCGCGAGTTCCACCTCGCGCCTCCCTCTCTAGACAGGCGCTGGGAGAAAGAACCG	230
Qy	887	cttgagaact-gggctctgtgccagcgcgaggtgcagg	924
Db	231	CTCCCAGATTCCGGCAATTTCCGCCGGCTCGAGGTGCAGG	269

RESULT 15

RESULT 13  
US-08-586-165-1/c  
; Sequence 1, Application US/08586165  
; Patent No. 6054298  
; GENERAL INFORMATION:  
; APPLICANT: Laufer, Edward M.  
; APPLICANT: Orozco, Olivia E.

```

1 APPLICANT: Tabin, Clifford J.
2 TITLE OF INVENTION: Fringe Proteins and Pattern Formation
3 NUMBER OF SEQUENCES: 9
4 CORRESPONDENCE ADDRESS:
5 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
6 STREET: Two Militia Drive
7 CITY: Lexington
8 STATE: Massachusetts
9 COUNTRY: US
10 ZIP: 02173
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/586,165
19 FILING DATE: 16-JAN-1996
20 CLASSIFICATION: 800
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Granahan, Patricia
23 REGISTRATION NUMBER: 32,227
24 REFERENCE/DOCKET NUMBER: H095-05
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (617) 861-6240
27 TELEFAX: (617) 861-9540
28 INFORMATION FOR SEQ ID NO: 1:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 1120 base pairs
31 TYPE: nucleic acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34 FEATURE:
35 NAME/KEY: CDS
36 LOCATION: join(1..240, 244..474, 478..531, 538..579, 58
37 LOCATION: .678, 682..687, 691..807, 811..843, 850..870
38 LOCATION: 874..990, 994..1056, 1060..1083, 1087..1104,
39 LOCATION: .1119)
40 US-08-586-165-1

```

Query Match 4.3%; Score 40; DB 3; Length 1120;  
Best Local Similarity 51.7%; Pred. No. 0.076;

[illegible]

Search completed: June 21, 2001, 18:21:46  
Job time: 17499 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 15:34:03 ; Search time 4579.66 Seconds  
(without alignments)  
1907.224 Million cell updates/sec

Title: US-09-445-201-1\_COPY\_6036\_6959  
Perfect score: 924  
Sequence: 1 gaagttcacacacgaaatgt.....tgccacgcgaggtgcagg 924

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_estl1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est33:\*  
111: gb\_est34:\*  
112: gb\_est35:\*  
113: gb\_est36:\*  
114: gb\_est37:\*  
115: gb\_est38:\*  
116: gb\_est39:\*

117: gb\_est48:\*  
 118: gb\_est49:\*  
 119: gb\_est50:\*  
 120: gb\_est51:\*  
 121: gb\_est52:\*  
 122: gb\_est53:\*  
 123: gb\_est54:\*  
 124: gb\_est55:\*  
 125: gb\_est56:\*  
 126: gb\_est57:\*  
 127: gb\_est58:\*  
 128: gb\_est59:\*  
 129: gb\_est60:\*  
 130: gb\_est61:\*  
 131: gb\_est62:\*  
 132: gb\_est63:\*  
 133: gb\_est64:\*  
 134: gb\_est65:\*  
 135: gb\_est66:\*  
 136: gb\_est67:\*  
 137: gb\_est68:\*  
 138: gb\_est69:\*  
 139: gb\_est70:\*  
 140: gb\_est71:\*  
 141: gb\_est72:\*  
 142: gb\_est73:\*  
 143: gb\_est74:\*  
 144: gb\_est75:\*  
 145: gb\_est76:\*  
 146: gb\_est77:\*  
 147: gb\_est78:\*  
 148: gb\_est79:\*  
 149: gb\_est80:\*  
 150: gb\_est81:\*  
 151: gb\_est82:\*  
 152: gb\_est83:\*  
 153: gb\_est84:\*  
 154: gb\_est85:\*  
 155: gb\_est86:\*  
 156: gb\_est87:\*  
 157: gb\_est88:\*  
 158: gb\_est89:\*  
 159: gb\_est90:\*  
 160: gb\_est91:\*  
 161: gb\_est92:\*  
 162: gb\_est93:\*  
 163: gb\_est94:\*  
 164: gb\_est95:\*  
 165: gb\_est96:\*  
 166: gb\_est97:\*  
 167: gb\_est98:\*  
 168: gb\_est99:\*  
 169: gb\_est100:\*  
 170: gb\_est101:\*  
 171: gb\_est102:\*  
 172: gb\_est103:\*  
 173: gb\_est104:\*  
 174: gb\_est105:\*  
 175: gb\_est106:\*  
 176: gb\_est107:\*  
 177: gb\_est108:\*  
 178: gb\_est109:\*  
 179: gb\_est110:\*  
 180: gb\_est111:\*  
 181: gb\_est112:\*  
 182: gb\_est113:\*  
 183: gb\_est114:\*  
 184: gb\_est115:\*  
 185: gb\_est116:\*  
 186: gb\_est117:\*  
 187: gb\_est118:\*  
 188: gb\_est119:\*  
 189: gb\_est120:\*

190: gb\_est121:\*  
 191: gb\_est122:\*  
 192: gb\_est123:\*  
 193: gb\_est124:\*  
 194: gb\_est125:\*  
 195: gb\_est126:\*  
 196: gb\_est127:\*  
 197: gb\_est128:\*  
 198: gb\_est129:\*  
 199: gb\_est130:\*  
 200: gb\_est131:\*  
 201: gb\_est132:\*  
 202: gb\_est133:\*  
 203: gb\_est134:\*  
 204: gb\_est135:\*  
 205: gb\_est136:\*  
 206: gb\_est137:\*  
 207: gb\_est138:\*  
 208: gb\_est139:\*  
 209: gb\_est140:\*  
 210: gb\_est141:\*  
 211: gb\_est142:\*  
 212: gb\_est143:\*  
 213: gb\_est144:\*  
 214: gb\_est145:\*  
 215: gb\_est146:\*  
 216: gb\_est147:\*  
 217: gb\_est148:\*  
 218: gb\_est149:\*  
 219: gb\_est150:\*  
 220: gb\_est151:\*  
 221: gb\_est152:\*  
 222: gb\_est153:\*  
 223: gb\_est154:\*  
 224: gb\_est155:\*  
 225: gb\_est156:\*  
 226: gb\_est157:\*  
 227: gb\_est158:\*  
 228: gb\_est159:\*  
 229: gb\_est160:\*  
 230: gb\_est161:\*  
 231: gb\_est162:\*  
 232: gb\_est163:\*  
 233: gb\_est164:\*  
 234: gb\_est165:\*  
 235: gb\_est166:\*  
 236: gb\_est167:\*  
 237: gb\_est168:\*  
 238: gb\_est169:\*  
 239: gb\_est170:\*  
 240: gb\_est171:\*  
 241: gb\_est172:\*  
 242: gb\_est173:\*  
 243: gb\_est174:\*  
 244: gb\_est175:\*  
 245: gb\_est176:\*  
 246: gb\_est177:\*  
 247: gb\_est178:\*  
 248: gb\_est179:\*  
 249: gb\_est180:\*  
 250: gb\_est181:\*  
 251: gb\_est182:\*  
 252: gb\_est183:\*  
 253: gb\_est184:\*  
 254: gb\_est185:\*  
 255: gb\_est186:\*  
 256: gb\_est187:\*  
 257: gb\_est188:\*  
 258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	211	22.8	763	169	BF784985
2	157.4	17.0	192	161	BB597428
3	151.2	16.4	207	150	BF543874
4	146.2	15.8	193	161	BB593727
5	79.2	8.6	335	106	AL589555
6	62	6.7	925	219	CNS0091P
7	55.8	6.0	925	219	CNS0091P
8	52.6	5.7	645	219	CNS01213
9	50.8	5.5	935	219	CNS006XK
10	48	5.2	920	219	AG013041
11	48	5.2	1071	219	CNS006MY
12	47.4	5.1	839	219	CNS004NB
13	46.4	5.0	692	219	CNS007WH
14	46.2	5.0	1101	219	CNS012PK
15	44.8	4.8	935	219	CNS006XK
16	44.2	4.8	1201	219	CNS01523
17	44	4.7	844	219	CNS0052P
18	43.8	4.7	930	239	AZ192144
19	42.6	4.6	932	219	CNS0072Q
20	42.4	4.6	1101	219	CNS017SY
21	42.2	4.6	402	14	AA962730
22	41.8	4.5	243	235	AQ937916
23	41.8	4.5	839	219	CNS004NB
24	41.6	4.5	932	219	CNS0072Q
25	41.6	4.5	1009	219	CNS010EW
26	41.2	4.5	1203	219	CNS015Y4
27	40.8	4.4	874	172	BF983618
28	40.6	4.4	456	110	AW002648
29	40.4	4.4	510	235	AQ938393
30	40.2	4.4	844	219	CNS0052P
31	40	4.3	894	219	CNS01591
32	39.4	4.3	948	145	BF203166
33	39.4	4.3	1101	219	CNS012S8
34	39.2	4.2	788	170	BF860542
35	39.2	4.2	910	219	CNS006ON
36	39	4.2	397	174	BG148343
37	39	4.2	584	141	BE912455
38	39	4.2	618	150	BF580660
39	39	4.2	701	138	BE648283
40	38.8	4.2	426	219	CNS016VV
41	38.8	4.2	692	219	CNS007WH
42	38.8	4.2	1201	219	CNS014BJ
43	38.6	4.2	926	239	AZ186391
44	38.6	4.2	954	145	BF143441
45	38.6	4.2	1101	219	CNS001XJ

## ALIGNMENTS

RESULT 1  
 BF784985 763 bp mRNA EST 12-JAN-2001  
 LOCUS 602110935F1 NCI\_CGAP\_Kid14 Mus musculus CDNA clone IMAGE:4238984  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BF784985  
 VERSION BF784985  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 763)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.

Email: [ccapbs-re@mail.nih.gov](mailto:ccapbs-re@mail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM9851 row: f column: 09  
 High quality sequence stop: 657.  
 Location/Qualifiers

source

1. 763  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4238984"  
 /lab\_host="NCI\_CGAP\_Kid14"  
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library. |"  
 BASE COUNT 156 a 218 c 223 g 166 t  
 ORIGIN

Query Match 22.8%; Score 211; DB 169; Length 763;  
 Best Local Similarity 98.5%; Pred. No. 3.6e-48;  
 Matches 255; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

QY 666 t g t t t c c t t a g a c g c g c a c c t a c c c g c a g a a g c c a g a c t g t g t c c 725  
 Db 1 T G T G T T T C C T A G A C G C G C G G A - C G T A C C C G C A G A C T G A A G C C C A C A C T G T G T C C 59  
 QY 726 c g c g c g g g a t a c c t g g t g a c c c g a t t c c g c g a c c g t g c a g c g g g t g g a g 785  
 Db 60 C G C A G - C G G G A T A C C T G G T G A C C C G A T T C C G G G A C A C C G T G C A G C C G C G G T G G A G 118  
 QY 786 c c a g g c g c g g t c c c c g c g c t c c c g t t t g g t g c g g g g g c g a t a c c g c t c 845  
 Db 119 C C A - G G C C C G G T C C C C G C C T C T C C C C G G T C T T G C G T C G G G G G G C A T A C C G C T C 177  
 QY 846 t g t g a c t t c t t g c g g c a g g g a g a g a g a g t g t g c c t g a g a a c t g g g c t c t g t 905  
 Db 178 T G T G A C T T C T T G C G G G C C A - G G A G G A G A G A G A G T C T G C C T G A G A A C T G G G C T C T G T 236  
 QY 906 g c c a g c g c g a g t g c a g g 924  
 Db 237 G C C A G C G C G A G T G C A G G 255

RESULT 2

BB597428 192 bp mRNA EST 01-DEC-2000  
 LOCUS BB597428 RIKEN full-length enriched, 12 days embryo spinal cord Mus  
 DEFINITION musculus cDNA clone C530002A01 5', mRNA sequence.  
 ACCESSION BB597428  
 VERSION BB597428  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 192)

Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,  
 Hanagaki, T., Hayata, S., Hirozane, T., Hirozane, T., Hodozane, Y.,  
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno  
 , H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K.,  
 Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,  
 Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, K., Shinagawa, A.,  
 Shiraki, T., Soabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka  
 , T., Toya, T., Watanishi, A., Yamamura, T., Yasunishi, A., Yoshida, K.,  
 Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.



Query Match	16.4%	Score 151.2	DB 150	Length 207
Best Local Similarity	88.0%	Pred. No. 1.4e-31		
Matches 176	Conservative 0	Mismatches 23	Indels 1	Gaps 1
Qy 654	aagagagtaagctgtgttctcttagatcgcgagccgactaccgagcaggaactgaaagcc	713		
Db 8	AAGAGAGTAGCTGTGTCTCTTATAGTCGCCGGAACCGTACCCCGCAGCACTGAAAGCC	67		
Qy 714	cagactgtgtccccgcagcgggataaacctggtgacccgattcccgcgacaccgcctgcag	773		
Db 68	CAGATTGTGTCTGCAGCGGGATAAACCCTGCTGACCGGTCTCCGCCGACACTGCTGCAC	127		
Qy 774	ccgcgctggagccagggcgcc-ggtgcccgcgctctccccggtctctgctgcgtgcgggg	832		
Db 128	CCGCGGCTGGAGCCAGGCGCGGCTGCGCTTCCAGGTTTTCGCACTGCAAGGAG	187		
Qy 833	cgcataccgctctgtgact	852		
Db 188	CCGCTAAGCCTTCTGTGACT	207		
RESULT 4				
LOCUS BB593727	193 bp	mrna	EST	30-NOV-2000
DEFINITION BB593727	RIKEN full-length enriched, 4 days neonate male adipose			
	Mus musculus cDNA clone B430204G23 5', mRNA sequence.			
ACCESSION BB593727				
VERSION BB593727.1	GI:11490329			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	1 (bases 1 to 193)			
	Alizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,F., Hirozane,T., Hodojima,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Taya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.			
TITLE	RIKEN Mouse ESTs (Alizawa,K. et al. 2000)			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Yoshihide Hayashizaki Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel: +81-298-36-9013 Fax: +81-298-36-9098 Email: genome-res@rtc.riken.go.jp, URL: http://genome.rtc.riken.go.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site ( <a href="http://genome.rtc.riken.go.jp">http://genome.rtc.riken.go.jp</a> ) for further details.			
FEATURES	Location/Qualifiers			
source	1..193			
	/organism="Mus musculus"			

/db_xref="taxon:10090"				
/clone="B430204G23"				
/clone_lib="RIKEN full-length enriched, 4 days neonate male adipose"				
/sex="male"				
/tissue_type="adipose"				
/dev_stage="4 days neonate"				
/lab_host="Dhl08"				
/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGATCCAAAGACTCTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 229.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTAAATAAATTAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLIC I."				
BASE COUNT	38 a	44 c	82 g	29 t
ORIGIN				
Query Match	15.8%;	Score 146.2;	DB 161;	Length 193;
Best Local Similarity	85.3%;	Pred. No. 3.5e-	30;	
Matches 163;	Conservative	0;	Mismatches 28;	Indels 0;
Gaps	0;			
Qy	638	gagtcctcaggaccccccaagagatgaactgtgttccttagatcgcgagaccgtacc	697	
Db	3	GAGTCCTCAGAACCCCAAGAGAGTAGTGTTCTTTAGATCGGGGGGAACGGTACC	62	
Qy	698	ggcaggactgaagcccgagactgtgtcccgcagccggataaacctggctgaccgcgatcc	757	
Db	63	GGCAGGAGTGAAGCGCAGAAATGTGCCGGAGCGGGAAAAATCTGGCTGAACCGAATCC	122	
Qy	758	gcggacaccgctgcagccgcggtggagcacaggcggcggtgcccgctctcccccggt	817	
Db	123	GGGAAACCGTTGCAGCCGGGGGTGGAGCAGGCGCGCGGTCTCCCCTCCGCT	182	
Qy	818	cttcgcgtcgq	828	
Db	183	CTGGGGGTGGG	193	
RESULT	5			
AL589555				
LOCUS	AL589555	335 bp	mRNA	EST
DEFINITION	DKFZp451M0219_r1_451 (synonym: hlcc1) spinal cord Homo sapiens cDNA clone DKFZp451M0219 5', mRNA sequence.			07-MAR-2001
ACCESSION	AL589555			
VERSION	AL589555.1	GI:13243327		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Bloeker,H., Boecher,M., Brandt,P., Mwes,W., Weill,B. and Wiemann,S.			
TITLE	EST (Bloeker,H., Boecher,M., Brandt,P., Mwes,H.W., Weill,B. and Wiemann,S.)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Bloecker H MPS Am Klopferspitz 18a D-82152 Martinsried, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer			





REFERENCE	1 (bases 1 to 935)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 0006 EVRY cedex - FRANCE (E-mail : <a href="mailto:segr@genoscope.cns.fr">segr@genoscope.cns.fr</a> ) - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila/bac.htm">http://bacpac.med.buffalo.edu/drosophila/bac.htm</a> .

BASE COUNT	ORIGIN	257 a	170 c	162 g	96 t	250 others
Query Match	5.5%	Score 50.8;	DB 219;	Length 935;		
Best Local Similarity	30.3%;	Pred. NO. 0.002;				
Matches 125;	Conservative 100;	Mismatches 185;	Indels 2;	gaps 2;		

SOURCE Homo sapiens DNA, Clone:72018Spn42.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 920)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.  
TITLE Homo sapiens genomic DNA, chromosome 21q  
JOURNAL Published Only in DataBase (1998) In press

REFERENCE 2 (bases 1 to 920)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (25-AUG-1998) to the DDBJ/EMBL/GenBank databases.  
Masahira Hattori, Kitasato University, Department of Science, JST  
Sequencing Laboratory, Kitasato 1-15-1, Sagamihara 228, Japan  
(E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,  
Fax:0427-78-9561)

COMMENT On Feb 5, 1999 this sequence version replaced gi:2913972.  
AG005396; Submitted (26-Feb-1998).

FEATURES  
source  
1..920  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/clone="72018Spn42"  
/map="21q"

BASE COUNT 16 a 154 c 414 g 12 t 324 others

ORIGIN  
Query Match 5.2%; Score 48; DB 219; Length 920; -  
Best Local Similarity 36.0%; Pred. No. 0.012;  
Matches 123; Conservative 0; Mismatches 219; Indels 0; Gaps 0;  
QY 518 cgggaacccaaacctggtatccagtggggcgtgcgcgacgcaggagtgccccacc 577  
Db 494 CNGGNCNCCCGCCCGCTGNNCGNCGNCGNGGAATNGCGGCCCGACNCCCAGTG 435  
QY 578 ctcccggtaatgacccgcgcgccttcctgctagtgtgtagccgcgctctcttctgcct 637  
Db 434 CCCCCCGGNNCCCGCGNCCGGCCGCTGATCCCGNNNNNCGNCCGNC 375  
QY 638 gactctcaggaccccaagagtagctgtgttctctagatcgcgagaccgtacc 697  
Db 374 NCCNCCNNGGNGCNAGCNNTGCCGNNCCGNNCCCTCANNCCCCGGGCCCGAGCT 315  
QY 698 ggcaggactgaagccagactgtgtccgcagccggataaactggctgaccgattcc 757  
Db 314 GCNCCGCCNNNNNCCGCGNCCNCCNCCGCGCCNCCNCGCTCGAGNCCN 255  
QY 758 gcggacacgcgtcagccgcggtgtagacagcgccggtgcgccgcgctctcccggt 817  
Db 254 CNGGNCNCCGCGCCCTGNGGNTNCCNCGAGGCGCCNCGNCCNCCNCCGNCN 195  
QY 818 ctgcgcgtcgggggcgcataccgcctctgtgactctttgc 859  
Db 194 CNGGNNCCNCCGCGCCCGCCNCCGNNCCGCAATNCCNCC 153

RESULT 11  
LOCUS CNS000EMY 1071 bp DNA GSS 04-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR29L07 of RFL1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL069395  
VERSION AL069395.1 GI:4949538  
KEYWORDS gss.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1071)





```

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR16H04 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL050923
VERSION AL050923.1 GI:4932977
KEYWORDS fruit fly,
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 692)
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aron Mammosier in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR16H04"
                     /note="end : TET3"
BASE COUNT    118 a   69 c   104 t   333 others
ORIGIN

Query Match      5.0%; Score 46.4; DB 219; Length 692;
Best Local Similarity 17.4%; Pred. No. 0.033;
Matches 56; Conservative 139; Mismatches 126; Indels 0; Gaps 0;

Qy 425 ccaccagaaacccagctgctccagattgtctcagatgcgacttgcgcgcgcgcacag 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 SNVAGSVRVVSVVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSV 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 485 ttccggggttagtgggagtggtggtggaacccggaacccaaacctgttccagtg 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 SRMSGSSARVCGSGVSGMSGSSGSSAGRVAAAGSAGAGSGVSGSSGBCACVSVMS 489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 545 ggggggtggtggcagcagcaggttcccccacccctcccgttaagaccaccccccattc 604
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 SSRVAVSMGSGVCGCVVVMVCMCRVMMMKCCCVCGVSSCSACCCCVVMAACRCVAV 549
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 605 gctadgtgtagcggcgctctcttttgcctgtagctcagcagcccaagagagtaag 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 SCRAGAACMAASVAVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSV 609
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 665 ctgttttcttagtcgcgcgcgcgcgtaccgccgagactgaaagccagactgtgtc 724
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 CACASRVAVCCCMSCAMASVSGSGARGVRSASSSSCSASMSMCMCAVAMSCVAVS 669
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 725 ccgcagccgggataacctggc 745
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 670 SRSSVAVGVSVMMSVSAVC 690
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14

```

```

CNS012PK      1101 bp      DNA      GSS      26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN08K01 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL101858
VERSION AL101858.1 GI:5613469
KEYWORDS fruit fly,
SOURCE Plasmid Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelOBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN08K01"
                     /note="end : T7"
BASE COUNT    80 a   251 c   335 g   63 t   372 others
ORIGIN

Query Match      5.0%; Score 46.2; DB 219; Length 1101;
Best Local Similarity 34.8%; Pred. No. 0.041;
Matches 81; Conservative 1; Mismatches 151; Indels 0; Gaps 0;

Qy 471 gcccccccgcacagttcccggttagtggggagtg999cggtgggaaacccggaaccccaaa 530
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 GCTCCCCGNCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 531 cctggtatccagtggtggggtggtggtggtggtggtggtggtggtggtggtggtggtggt 590
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 GNGCNCGACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 591 ccccgccccattcgctagtgtagcggcgctctcttcttgcctgagctcagcagac 650
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 GGCCGCGCNCNNNGCGNNNNNNNNNGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 581
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 651 cccaagagagtaagctgtgttcttccttagatgcgcgcgcgcctaccgcagcagg 703
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 NGGCGNANNNGCGCNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGG 634
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
CNS006XK/c     935 bp      DNA      GSS      03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL066051
VERSION AL066051.1 GI:4945019
KEYWORDS fruit fly,
SOURCE Plasmid Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 935)

COMMENT

Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> the BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammos in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Location/Qualifiers  
1..935  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACRI14N09"  
/note="end : T7"

BASE COUNT 257 a 170 c 162 g 96 t 250 others  
ORIGIN

Query Match

Best Local Similarity 4.8%; Score 44.8; DB 219; Length 935;  
Matches 108; Conservative 94; Mismatches 179; Indels 1; Gaps 1;  
Qy 544 ggggggctggccgacgagggagtcaccccccctcccggtatgaccccgccccatt 603  
Db 914 SGGSGSGGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 855  
Qy 604 cgctagtgtgtagcggcgctctcttctgcccctgagtcctcagagccccaagagataa 663  
Db 854 CGCGCGSSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 795  
Qy 664 gctgttttcttagatcgccgacgacctaccggcaggactgaaagccagactgtgt 723  
Db 794 GCCSGCGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735  
Qy 724 cccgcagcgg-ggataacctggtgacccggtatcccgagacacgctgacgcggtg 782  
Db 734 SSCSSGCGSGSGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675  
Qy 783 gagccagggcgccgtgcccccgctctcccggctcttgcgctgcggggcgccataccg 842  
Db 674 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 615  
Qy 843 ctctgtgactcttgcggccagggagagagagagagagagagagagagagagagag 902  
Db 614 CCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 555  
Qy 903 tgtgccagcgaggtgcagg 924  
Db 554 GSAGSSGRCGACGSGGSSBSS 533

Search completed: June 21, 2001, 15:34:07  
Job time: 7651 sec

